

Spanish universities and the obstacles to development

A viable science base requires a commitment to excellence and imagination that is incompatible with rigidity and cronyism. Spain needs to absorb this lesson if it is to flourish scientifically and economically.

Few countries are as well placed to turn modern science to their economic and social advantage as Spain. On the one hand, Spain's membership of the European Union should, in principle, allow it to tap into the best that Europe's research laboratories can offer. On the other, it enjoys the long-term prospect of expanding markets in Latin America. Yet a traditionally heavy-handed bureaucracy, the legacy of scientific stagnation during the Franco years, and the side-effects of a strong libertarian backlash when the Franco yoke was lifted, still threaten to hold the country back from exploiting this opportunity to the full. The consequent obstacles to progress are damagingly frustrating to Spain's best scientists who, despite this, continue to support their country's scientific and economic development.

Take the issue of university appointments. Current laws on universities, introduced in 1983, rightly give such institutions considerable freedom in appointing tenured faculty members (previously, these had been appointed centrally). But there is much evidence that the procedure under which such appointments are made, in particular the make-up of five-member appointments boards, two members of which come from the university involved, can remain excessively influenced by non-academic considerations. The concern is that the autonomy of universities has allowed the appointments process to become dominated by mutual self-interest operating through self-sustaining social networks — in other words, 'cronyism'.

The issue has been highlighted by charges brought by one astrophysicist against the University of Salamanca, who had returned to Spain — ironically in response to a government appeal — after working for eight years in Germany (see page 712). The astrophysicist claims that he was inappropriately, and illegally, rejected for a faculty appointment, despite the superiority of his scientific qualifications. In drawing attention to this case, *Nature* is not seeking to make any judgement on the validity of the university's decision to award two different posts, for which the astrophysicist had applied, to internal candidates. Nor, indeed, is the case necessarily different from

thousands of similar but less formal complaints that cronyism overshadows scientific performance in the appointments process.

The situation in Spain is by no means unique, but that should not make it any more acceptable. To its credit, the current administration in Madrid has shown itself aware of the problem. Proposed changes to university laws currently under consideration would alter the make-up of appointments committees, requiring at least four of the five members to come from other universities (selected at random). Such a move would be a welcome challenge to the 'cronyist' temptations.

But the problem goes deeper than voting procedures. Social networks will always find ways to flourish, whatever is done to limit their effectiveness. And changing the appointments process on its own will not address the second major problem facing universities in Spain (and elsewhere): the intellectual sclerosis created by a broad system of tenure that requires little accountability from researchers, after they have crossed initial hurdles, for the rest of their working lives. The trick is to find a way of improving the situation that does not undermine the commitment to long-term goals on which the health of a growing science base depends.

The two problems need to be addressed together. Again, this appears to be recognized by many of those government officials who are responsible for the health of the nation's research base, for example in the proposal that universities provide four-year teaching contracts. It seems to be less accepted in wider political circles, where those who enjoy the privileges and security of the current arrangements — and who would be directly threatened by any significant changes — can often find substantial support. But change is essential if Spain is to achieve its full scientific and technological potential. Already some institutions, such as the nascent National Centre for Cancer Research in Madrid, are demonstrating that alternative arrangements are possible; hopefully, a high scientific output from such 'experiments' will show that they are also desirable. Spain's politicians should give them close attention. □

Kemp's cultures

A series that ends this week should inspire not only appreciation of images but also a desire to communicate.

Irresistibly and in occasionally peremptory fashion, science has, in recent years, embedded itself ever deeper into everyday culture. One can reasonably assume that this infiltration will grow as science significantly extends and deepens our understanding of just what we are.

Visually, that cultural permeation can be witnessed at several levels. Citizens are surrounded by images, often taken out of scientific context, with values ranging from the ephemeral to the icon with an impact that resonates for decades. At a deeper level, scientific ideas and perspectives are absorbed by artists and significantly enrich their portrayals or creative extensions of the world.

For over a year, Martin Kemp has treated readers of *Nature*, week by week, to a seemingly endless celebration of diverse and often beautiful artistic and scientific images and the (respectively) scientific and artistic impulses that can be traced within them. But the end has finally come (page 727). To judge by the author's electronic mailbox, readers have been much stimulated. They will be pleased to know that he will continue to contribute regularly, albeit less frequently, in future. Meanwhile, researchers might consider the ever more important need to communicate their science not only to their peers but also in ways that others can appreciate and that artists in particular, occasionally but valuably, can creatively absorb. □



World Bank backs Third World centres of excellence plan

[WASHINGTON] A global chain of so-called Millennium Institutes, acting as scientific centres of excellence in developing countries to galvanize a rapid increase in their scientific and technical strength, is being planned by the World Bank, private foundations and several governments.

The World Bank could grant the first loan, of around US\$5 million, to establish prototypes for such institutes in Chile as early as next month, according to bank officials. The initiative has the strong personal support of Eduardo Frei, the president of Chile, and James Wolfensohn, the bank's president.

The institutes will be characterized by the fact that leading scientists from outside the host country will select their directors and review their performance, that the focus of research will be of direct relevance to that country's economic and social needs, and that there will be frequent exchange of students and researchers with institutes abroad.

Frei has been discussing ideas for new scientific collaboration with other Latin American leaders for more than a year (see *Nature* 391, 524–525; 1998). Wolfensohn, who is on the board of directors at the Institute for Advanced Study (IAS) at Princeton, New Jersey, and has been instrumental in focusing the bank on the 'knowledge gap' between industrialized and developing countries (see *Nature* 395, 529; 1998), badly wants to bolster science in the bank's client countries.

Wolfensohn asked IAS director Philip Griffiths to investigate the possibility of new types of institutes to do this. Earlier this year Griffiths met Claudio Teitelboim, Frei's science adviser. Their discussion led to the convergence of the two initiatives at a special meeting in June in Santiago, attended by scientific leaders from Latin America, the United States and Europe.

The meeting resulted in a concept for the Millennium Institutes which, supporters



Supporting the best: Chile's president Eduardo Frei (right) and science adviser Claudio Teitelboim meet physicist Stephen Hawking.

believe, will enable a small number of excellent researchers in developing countries to break free from the constraints on first-class international research there.

They believe this can be achieved by investing a relatively small amount of money, directed at exceptional talent with guaranteed long-term support and subject to review by leading scientists from other countries.

"We want to have a research group working happily and feeling that they are not hindered by being in Chile rather than in Paris, New York or London," says Teitelboim.

Teitelboim says Chile's proposal is "a Chilean initiative that has the support of the World Bank". But bank officials and other advocates of the Millennium Institutes expect it to be rapidly followed by similar institutes elsewhere in Latin America, and that if successful, the concept will quickly spread to other regions of the world, including East Asia and Africa.

Bruce Alberts, the president of the US National Academy of Sciences, who attended the Santiago meeting, says that international peer review of the Millennium Institutes will

"keep out the politics that dominates science in developing countries".

He adds that the institutes will seek to transform the public image of science in these countries, where it is sometimes viewed as isolated from and subservient to economic and societal needs. "They'll need to demonstrate value beyond just their academic value," he says. "We want them to help bring science into the public consciousness." The institutes will "be embedded in the university system" of their host countries, where their staff will be expected to teach.

Lauritz Holm-Nielsen, a senior science and technology specialist at the World Bank who runs the initiative there, says its rate of expansion will depend on the interest expressed by the bank's client countries. "We're trying at first to work with Chile, Argentina, Brazil and Colombia," he says.

Holm-Nielsen adds that the institutes will not adhere to any particular model or structure, but will instead focus on supporting the best people, using "a variety of models" applicable to individual countries. The bank will expect host governments to adopt a policy of long-term support for the institutes.

It will rapidly consider small "learning and innovation" loans, such as that requested by Chile, to start prototypes, and then larger loans for the long-term operation of the institutes. One researcher familiar with the programme expects Chile to request \$40–\$60 million to operate its first batch of institutes.

Griffiths says several private foundations, including the Carnegie Corporation and the Fulbright programme, are involved in the discussions on the initiative or have already pledged grants to it. He hopes that around six prototype institutes will be established within a year.

For Chile, according to Teitelboim, "the key thing is to get the prototype to fly". He envisages "no more than three" prototype institutes in Chile, together with "a larger number of nuclei", where a single investigator will be supported, and which may later evolve into full institutes. A third element of funding will support visiting students and researchers. "There will be a very small number of permanent people, and a large flow of people visiting," says Teitelboim.

The Chilean government will not decide the disciplines in which the institutes will specialize, nor their balance between pure and applied research. "To decide whether Chile should excel in this or that would defeat the purpose" of the initiative, Teitelboim says. "We should scout for talent, and then follow that talent."

Colin Macilwain

Germany's nuclear secrets can now be told

[MUNICH] Formerly secret documents relating to Germany's nuclear programme between 1938 and 1945 were transferred to Munich's Deutsches Museum last week. The documents include research reports and experimental protocols written by Germany's Nobel prizewinners Werner Heisenberg and Otto Hahn, as well as their correspondence with the government.

In 1938 Hahn, Lise Meitner and Fritz Strassmann discovered the first evidence of the products of uranium nuclear fission.

After the beginning of the Second World War, the Nazis increased support for the nuclear research programme. The documents show that until the early 1940s Germany and the United States were almost equally advanced in nuclear research.

Immediately after the war the Americans confiscated the documents, but they were returned to Germany in the 1970s and kept at the Karlsruhe National Research Centre. The museum will make the documents freely available to historians.

Quirin Schiermeier

Spanish university sued for favouritism...

[BARCELONA] The University of Salamanca in Spain is being taken to court by an astrophysicist who claims that a bias towards internal candidates denied him an associate professorship in its department of general and atmospheric physics.

A university panel has already rejected an appeal by the astrophysicist, Antonio Ferriz Mas. It argues that, despite having better scientific qualifications than the individual appointed, his speciality of fluid dynamics was "significantly different" from the department's requirements, namely atmospheric dynamics and thermodynamics.

But Ferriz Mas says the university has violated the principles of equality and meritocracy in assessing applicants for the post. University officials are not commenting publicly on the charges, which will probably take one to two years to reach court.

The administration of public universities in Spain has improved following reforms in 1983. But the selection process for professors, known as the '*concurso-oposición*', is still widely seen as a considerable barrier to the development of high quality research.

A contract researcher already employed in a department is frequently given a tenured position, for example, even if they are not the most scientifically qualified applicant.

Under the 1983 law, appointment boards must have five members, at least three of whom must vote for the approved candidate. Two members come from the department concerned, and three from other universities. But often the crucial third vote for a local candidate is said to be easily obtained.

Many concerned parties — including vice-chancellors and officials at the Ministry of Education and Culture — are now backing a change in the law to give each university department only one vote, that of chair of the appointment board (see below).

Ferriz Mas worked at the University of Freiburg in Germany for three years as a PhD fellow and five as a postdoctoral researcher.



Antonio Ferriz Mas: appealing rejection.

He returned to Spain in October 1994 under a programme sponsored by the government for 'reincorporating' researchers. A month later he applied for a post in earth physics, astronomy and astrophysics at the University of Salamanca, and spent six months preparing the teaching project on atmospheric dynamics and thermodynamics.

There were three candidates for the post. In the first part of the examination, based on research experience and proposed teaching project, the examiners gave Ferriz Mas three points. The local candidate, who was put forward by the department and was eventually awarded the post, received the top mark of five. The second exam, which involved giving a lecture, resulted in an equal score.

Ferriz Mas appealed against the decision. A year later, an appeal panel, chaired by vice-chancellor Ignacio Berdugo Gómez de la Torre, reported after taking advice from two foreign scientists. Both agreed that Ferriz Mas's work was "substantially more important" than that of the appointed candidate.

Despite this, Ferriz Mas's claim was rejected because the first part of the appointment process depended on the match between the position's 'teaching profile' and the research and teaching programme proposed by the candidates. The appeal panel acknowledged that the appointment board "should have operated with a higher diligence, as the 'fit' to the teaching profile must be clearly stated in preliminary information, as well as in the reports of the first exam".

The panel said it lacked the knowledge to judge how much the need to meet this 'teaching profile' should take precedence over the quality of prior research and teaching.

The appointment process had already been questioned by the third candidate for the post, Fernando Atrio Barandela, associate professor of theoretical physics at the University of Salamanca. He was also given three points in the first exam, following which he withdrew voluntarily.

Atrio Barandela asked one member of the appointment board the reason for his score, and says he was told it was not intended to reflect the relative merits of the candidates, but rather the committee's preferences.

A year later, Ferriz Mas applied for another post, an assistant professorship in the theoretical physics group, with an astrophysics teaching profile, at the university's Faculty of Sciences. The post involved instruction in using telescopes, but was given to a chemist.

Again Ferriz Mas appealed. This time the appeal panel, again chaired by Gómez de la Torre, accepted that the person appointed had a worse fit to the teaching profile than Ferriz Mas — but had substantially better research and teaching experience.

Ferriz Mas says that this second decision confirms that the local candidate was the predetermined choice for each post, and that there is often an unspoken agreement among the members of appointment boards.

In such cases, he argues, the deciding third voice seals a pact where board members return the favour in the future. He says that committee members have total impunity.

Many Spanish researchers are critical of the process. Benjamin Montesinos, director of the laboratory of space astrophysics and fundamental physics, an 'associated unit' of the Institute of Astrophysics of Andalusia (Higher Council of Scientific Research), says that a smokescreen of complex arguments is often used to justify appointing an internal candidate to a post.

He argues that examiners who assess applicants to jobs in universities should have no link with the department concerned.

Luis F. Rull, professor of physics at the University of Seville, says the argument used against Ferriz Mas — that fluid dynamics is a different discipline to atmospheric dynamics and thermodynamics — is false. Rull calls for common standards and wants a public ranking of disciplines, perhaps based on publication record, impact factor and citations.

Juan J. Manfredi, professor of mathematics at the University of Pittsburgh, says that many qualifications are used by appointment boards to disqualify candidates who may be better than local candidates.

He says that some positions define a teaching profile (instead of area of knowledge) so closely that only one candidate can meet the criteria. This process, he argues, goes directly against the cross-fertilization that research requires.

Xavier Bosch

... as new posts created in academic reforms

[MADRID] The Spanish government announced last week that it plans to introduce a new university position, based on a four-year contract, with the same responsibilities and salaries as existing tenured positions.

The new contracts form part of a set of university reforms to be debated in parliament next month. Also included is a proposal to reduce the number of local

university representatives on five-member appointments committees from two to one (see main story).

The government hopes that both moves, which have been approved by the secretary of state for universities, Manuel Jesús González, will increase the employment prospects for well qualified PhDs. They stress that winning one of the new teaching contracts

will depend on an independent national agency evaluating a candidate's research experience.

It says that the proposed changes to the legal framework governing universities — whose funding and administration is the responsibility of individual provinces — will increase the rigour and openness of the selection process and raise teaching standards.

Europe bids for molecular biology 'club'

[MUNICH] The European Molecular Biology Organization (EMBO), which promotes molecular biology through fellowships, conferences and workshops, is to launch a series of young investigator awards.

EMBO hopes that the five-year awards will be sufficiently prestigious to encourage the brightest young scientists to install themselves permanently in laboratories across Europe. Those eligible must be within eight years of completing their PhD.

The intention is that EMBO Research Scientists, as the recipients of the awards will be called, will form a 'club' of top-level molecular biologists who will maintain links with each other and with research group leaders at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, for at least the duration of the fellowships.

The governments of EMBO's 23 member states, which operate collectively as the European Molecular Biology Conference, last week agreed in principle to finance the scheme. If all goes smoothly, they will decide on its permanent level of financing in 2000.

EMBO is expected to launch the scheme by contributing ECU500,000 (US\$590,000)

of its current budget annually into a pool of funds, which is expected to be expanded by donations from other sources such as charities and foundations. Some member countries will also be expected to contribute to this pool, the largest contributions coming from the richer countries.

EMBO Research Scientists will be given a salary, the costs of running a small laboratory conducting research independent of the host institute's tenured scientists, and travel expenses associated with 'networking' — attending an annual meeting and visiting other laboratories in the 'club'.

Some candidates would already be receiving national young investigator awards. They would receive no support from EMBO for salaries and laboratory costs, but would receive full support for networking.

Any member state wishing to increase its number of EMBO Research Scientists in this category could do so by making a further payment. This could be used to top up the resources of all the EMBO Research Scientists in that country if the national young investigator award is judged to be significantly lower than the European average.

A second category would be selected in the first instance by EMBO. The relevant member state would agree the choice of candidate and pay a significant proportion of salary and running costs. The remainder would come from the EMBO awards pool.

"EMBO's selection committee will ensure the high standard of selected Research Scientists, and ensure that their quality is uniform throughout Europe," says Frank Gannon, director of EMBO. "I expect the initiative will become an important component of scientific life in Europe, one which identifies high-quality researchers at an early stage in their careers, and helps us to keep such individuals in Europe."

Maria Carmo Fonseca, a professor of cell and molecular biology at the University of Lisbon, who previously worked at EMBL, says the strong networking component will be particularly helpful in attracting back to "peripheral" countries, such as Portugal and Spain, young scientists who trained abroad.

The Portuguese Ministry of Research has already agreed to set up a young investigator award scheme that will allow it to exploit the EMBO scheme, she says.

Alison Abbott

South Korean researchers under fire for claims of human cloning

[TOKYO] A South Korean research team, which announced last week that it had cloned a human embryo, has been criticized by other researchers for refusing to disclose the evidence to back up its claim.

Indeed, many doubt whether the cloning experiment, claimed to be the first to have produced an embryo from a human somatic cell, has produced promising results.

Researchers from Kyunghee University Hospital in Seoul announced at a press conference that they had cultivated an early-stage embryo using an unfertilized egg and a somatic cell from a female patient who had been receiving infertility treatment there.

The team, led by Kim Seoung Bo and Lee Po Yon, professor and associate professor in obstetrics and gynaecology at Kyunghee University, injected nuclei from granulosa cells (differentiated cells that surround the oocyte) into enucleated oocytes.

Of six fertilized eggs, one divided to the four-cell stage. Researchers say they then destroyed this 'embryo' without implanting it into a human body. They say they adapted the technique used by researchers from the University of Hawaii, who have cloned several generations of mice using granulosa cells (see *Nature* 394, 369–373; 1998).

"The experiment was carried out purely for the purpose of medical research, and not for clinical purposes. The fertilized egg was not implanted in the uterus of the donor, as



Anti-cloning protest: environmental groups hit out against the Kyunghee human cloning experiments.

we have no intention of creating a cloned baby," said Lee at the press conference.

Lee says the experiment was conducted according to ethical rules on artificial fertilization laid down by the South Korean Medical Association in 1993, which prohibit the implantation of genetically engineered human embryos. The government is

compiling a draft bill on the regulation of human cloning research, which is expected to pass the congress early next year.

Although the research team has no plans for any further experiments until the new legislation is ready, it hopes to use the technique to develop cells and tissues for replacement therapy and to treat infertility.

But there is widespread concern among the public, even though the experiment stopped short of producing a full human clone. The Korean Federation for the Environment Movement, for example, says human cloning research, if misused, could result in a "great mishap to all mankind".

There was a more muted response from the scientific community in South Korea, with some doubting its success, as no evidence has been released.

"The lack of a scientific presentation leads us to think that their results are not 'hair raising' in terms of a scientific breakthrough," says Dae-Young Uhm, associate professor in physiology at Sungkyunkwan University. "The only notable social response was from the civil rights groups."

Kim refuses to release evidence of the experiment's success to avoid further controversy. "We have strong evidence, including photos, but we will not release it now as we believe our reputation has been damaged enough," he says.

Asako Saegusa

Britain embraces 'knowledge economy'

[LONDON] The British government has pledged to put the commercialization of scientific knowledge at the heart of its industrial policy. The move comes in an ambitious and wide-ranging series of initiatives announced last week, including a white paper on how science can enhance economic competitiveness.

But implementing the initiatives is likely to be controversial. Some ministers in the Labour government are concerned that some

proposals clash with the priorities of other government departments.

The white paper (policy document) includes a new £150 million (US\$252 million) national venture capital fund to help finance small businesses with the potential to grow, such as high-technology companies. It also includes a £20 million annual Higher Education Reach Out fund to reward universities in England that work with businesses.

Universities in Scotland will receive £34 million over three years.

Existing schemes encouraging academics to work with industry are also being expanded, while £75 million for equipment is being given under the Joint Research Equipment Initiative (see *Nature* **396**, 607; 1998).

Some of the thinking behind the white paper is borrowed from the United States. For example, the science minister, Lord David Sainsbury, is to coordinate a series of studies into the extent to which high-technology companies in Britain can benefit by being located in clusters such as those to be found in Silicon Valley or around Boston.

The white paper includes a government-sponsored review of whether publicly funded research establishments are making the best use of intellectual property rights to maximize the commercial returns from research.

Launching the initiatives, Peter Mandelson, Secretary of State for Trade and Industry, said that, as a high-wage economy with high land and transport costs and few raw materials, the United Kingdom's best hope of raising economic growth rates is by exploiting the potential of 'knowledge industries'.

Many see the white paper as a significant step in the Labour party's shift away from its traditional socialist roots to its present position as a party of the centre-left, comfortable both with increased public investment and free markets. This is a shift that Mandelson personally helped to engineer, and he has strong backing from Sainsbury.

In a newspaper article published the day after the white paper's launch, Mandelson wrote that he was a politician "confident that his white paper marks a turning point in ideology and policy for his party and his country". The message was plain, he wrote: "Labour has dumped its interventionist past."

This language, and the emphasis on relying on science to create wealth, concerns ministers with more traditional socialist leanings. These include Michael Meacher, the environment minister, and his boss, John Prescott, the deputy prime minister.

Mandelson and Sainsbury will face one of their first tests over the development of high-technology clusters. These businesses are likely to be subjected to rigorous and lengthy planning applications, particularly if they are to be built in or near rural areas.

Mandelson says he wants planning applications from high-technology industries to be dealt with such that "the national economic interest is taken into consideration". But the environment department, which oversees the planning system, is anxious that environmental considerations should not be downgraded, according to officials.

A key test of this is likely to be an application by the Wellcome Trust to build a science

Making a case in the corridors of power

[LONDON] Britain's corridors of power last week witnessed a rather unusual gathering of scientists and senior ministers who – in the company of some potted plants – assembled in 10 Downing Street to brief the prime minister, Tony Blair, on why science matters, the issues surrounding public investment in science and its policy implications.

The presentations were made in a series of tightly orchestrated, four-minute slots, delivered around the long, boat-shaped Cabinet table. The potted plants, reputedly the first to be placed squarely on the Cabinet table in front of the prime minister, arrived with Caroline Dean from the John Innes Centre in Norwich, who talked about her work on the genes in the plant *Arabidopsis*.

Dean was among those whose talk came under the heading, "A Taste of Science". Any resonance with the billing found on the back of pre-prepared supermarket meals – "A Taste of India" – was a reminder that for most government ministers, science might as well be a far-flung foreign land.

But the scientists present left the meeting rather pleased. "I felt that science had moved to the top of the agenda," says Paul Nurse, director-general of the Imperial Cancer Research Fund.

"I'm prepared to be as cynical as the next person, but I came away feeling fairly impressed and upbeat about

the whole thing," says Matthew Freeman of the Medical Research Council's Laboratory of Molecular Biology in Cambridge.

According to the reports of those present, ministers talked of the recent success of the Research Assessment Exercise conducted on British universities, and there was cautious discussion of a possible new white paper on science.

It also appeared clear that senior ministers were not confident that the recent announcement of significant additional funds for science (see *Nature* **394**, 209; 1998) was likely to be sufficient to silence scientists, and they were expecting further calls for money.

The scientists in turn delivered some important messages about the difficulties of finding adequate funding for their work. One of those present was Polina Bayvel of University College London, representing the country's electrical engineering departments.

Already backed by a prestigious Royal Society University Research Fellowship, Bayvel has raised £1.5 million (US\$2.5 million) for an optical communications laboratory – but the effort has involved writing more than 35 grant proposals over the past five years.

The need for long-term funding for good scientists was driven home by Matthew Freeman, who used his own highly successful laboratory as an example.

Despite being billed as a mix of young and old, and life and physical scientists, the meeting inevitably ruffled a few feathers. "The make-up of the meeting clearly reflected a life-sciences bias," said one physicist. "This is understandable, but what's worrying is there might be an ideology to fund life sciences more because they create more wealth."

For many, however, the surprise turn of the meeting was the widely distrusted trade and industry secretary Peter Mandelson, who, despite his reputation as the Labour party's 'spin-doctor', seems to have made something of a hit with the assembled academics. "I really liked Mandelson," said one. "I hadn't expected to, as his image is an alarming one. [But] he was intelligent and had a really sharp, black sense of humour."

Towards the end came discussion of how to improve science's poor public image. National Science Week, suggested one participant, needed some high-profile events – such as a minister bungee jumping off a tall building to demonstrate Newtonian dynamics and Hooke's law.

"Haven't you heard?" quipped Mandelson. "I'm doing that from the Millennium Dome". Such an event would certainly serve to attract visitors to the government's £758-million dome, which houses an exhibition celebrating the millennium.

Natasha Loder

park next to its genetics research centre at Hinxton Hall near Cambridge. The planning inspector has rejected the proposal. But this decision has yet to be endorsed by Prescott.

Another battle looms over biotechnology regulation. The environment department plans to include more public representatives on its committee of scientists that advises the government on the safety of proposed genetically modified crops.

It also wants applications to grow such crops to be seen by an additional ethics committee. But the Department of Trade and Industry (DTI) is likely to oppose this on behalf of industry, which fears that further regulations will be time-consuming and a threat to economic competitiveness.

This battle will be fought out in a forthcoming review of the structure of Britain's biotechnology regulatory system, also announced last week. The review, to be carried out by the Cabinet Office and the Office of Science and Technology, will include public consultation on the regulatory process.

It has been set up partly in response to the collapse in public confidence in government science advice during the crisis over bovine spongiform encephalopathy, and partly to address public concern that regulations on the planting of genetically modified crops do not adequately address safety issues.

One senior environment civil servant says the spectre of a possible recession is one reason that the DTI is keen to help set up knowledge-based companies. But she says her department will face severe public criticism if environmental and safety considerations are relaxed.

Ehsan Masood

Foresight initiative goes for competitiveness

[LONDON] The British government last week chose the occasion of the publication of its white paper on industrial competitiveness (see left) to launch the second phase of its Foresight exercise.

The Foresight initiative was launched by the previous Conservative government with the aim of stimulating scientists and industrialists to think about how science can be better targeted at creating wealth and improving the quality of life.

The Department for Education and Employment has written to the higher-education funding councils saying that it expects them to promote Foresight, and to take appropriate steps "to maximize the commercial exploitation of university research".

This is among the first signs that Foresight priorities, which have come to dominate the allocation of the annual £1 billion (US\$1.68 billion) research budget from



Admiring progress: Mandelson visits a cancer research lab.

the four scientific research councils, will also influence the education department's research allocations.

The decision to launch the new Foresight exercise illustrates the government's commitment to its usefulness as a device for structuring research policy. But there is some unease within the Office of Science and Technology over the 'spin' given to it by Peter Mandelson and Lord Sainsbury in their presentation of the launch.

Both ministers emphasized Foresight's potential contribution to

economic competitiveness. But the second phase had initially been designed under Mandelson's predecessor Margaret Beckett, to tone down this aspect and to increase the emphasis on finding ways of using research to raise the quality of life (*Nature* **393**, 8; 1998).

The emphasis is less on finding technologies that would benefit particular industrial sectors, and more on getting scientists and industrialists to focus on various interdisciplinary 'themes' such as ageing, healthcare and crime prevention.

E. M.

Canadian whistleblower row prompts broader code of conduct

[MONTREAL] The Medical Research Council of Canada (MRC) is to attempt to broaden the ethical code of conduct for research involving humans that it published in September (see *Nature* **395**, 420; 1998). The decision follows a dispute between a clinical researcher, the pharmaceutical company funding her research, and the hospital where the research took place.

At present, the MRC code covers only research funded by itself and the two other principal fund-granting agencies, the Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council. It now wants the code to cover all research involving humans, regardless of who funds it. This is becoming increasingly important as government funding for research is being replaced by funding from industry.

The move follows a request for help from Toronto's Hospital for Sick Children to Henry Friesen, president of the MRC, following the release of a report about the activities of Nancy Olivieri, a researcher who

had been carrying out clinical studies of the drug deferiprone in the treatment of thalassaemia.

When Olivieri went public with warnings that the drug could cause liver fibrosis, Apotex, the drug's manufacturer which was paying for the research, disagreed with her findings and threatened her with legal action because she had signed a confidentiality agreement.

Olivieri claims that the hospital refused her legal aid to defend herself, a charge the hospital denies. When her research colleagues backed her, a public furor erupted. Olivieri and her supporters called for an independent inquiry into the affair but the hospital refused, agreeing only to set up an investigation of hospital policies and practices in general.

The hospital later changed its mind, and agreed to set up an inquiry into the affair itself. But Olivieri and her supporters refused to participate, claiming that the panel leader, Arnold Naimark, professor of medicine and physiology at the University of

Manitoba, had previous links with Apotex funds and so was not impartial.

On 9 December, the hospital released the panel's report, which exonerated the hospital from improper conduct, and said that patient safety was not compromised and that there were no conflicts of interest. But the report acknowledged the need to improve some hospital policies.

The report also criticized Olivieri for failing to report her concern about liver toxicity promptly to its Research Ethics Board. But Olivieri calls the report a whitewash.

Olivieri says that she and her colleagues refused to participate in the inquiry because of conflicts of interest on Naimark's part that she says are a matter of public record. "All the people with intimate knowledge of what happened were never questioned," she said.

She and her colleagues are determined to get an independent investigation into the matter. Observers say the affair illustrates the dangers of increasing industrial support for research in Canada.

David Spurgeon

Senators champion research agencies in budget plea to Clinton

[WASHINGTON] Twenty-four US Senators have called for President Bill Clinton to include more money for research agencies in his budget request for fiscal year 2000, which he is expected to present to the Congress on 1 February next year.

The bipartisan group wrote to Clinton on 11 December, expressing "concerns" about the plans of the White House Office of Management and Budget, currently completing the budget request. Early in December, science lobbyists warned that the office was considering very small increases next year for science agencies, because they had fared so well in this year's budget.

The senators pointed out that the Senate has passed the Federal Research Investment Act — an advisory measure which has not been considered yet by the House of Representatives — which would double research spending over 12 years.

Europe's Mars mission heads for drawing board

[MUNICH] The Mars Express mission of the European Space Agency (ESA) looks almost

certain to fly following the agency's decision last week to begin a detailed design phase. The mission is scheduled for launch in 2003.

ESA has been encouraged by the agreement of its council last week to provide an extra ECU3 million (US\$3.5 million) for the 1999 space science budget by reorganizing internal funds. But the council deferred to next May's ESA ministerial meeting the decision on the total funding for the mandatory science programme, which has been eroded over the past three years.

Cesarsky tipped to head Euro observatory

[MUNICH] French astrophysicist Catherine Cesarsky (below) is being hotly tipped as the next director of the European Southern Observatory in Garching, near Munich, to take over from Riccardo Giacconi next year.

Cesarsky is director of fundamental science at France's atomic energy research organization (CEA), and principal investigator of one of the main scientific instruments that flew on the European Space Agency's Infrared Space Observatory.



Another record year for global temperatures

[WASHINGTON] Average global temperatures were higher this year than in any year since record keeping began in 1860, the World Meteorological Organization reported last week. Based on observations up to mid-December, the organization estimates that the global mean surface temperature for 1998 — as measured by a network of ships, buoys and land-based weather stations — will be 0.58 °C above the average temperature for the years 1961 to 1990.

The previous warmest year, 1997, was 0.43 °C above the recent average. The 1990s have seen seven of the ten warmest years since the mid-nineteenth century.

'Water with a memory' protocols go public

[PARIS] Scientists have been offered an opportunity to attempt to reproduce the experiments of French scientist Jacques Benveniste, who claims to have shown that water has 'memory', and to be able to transmit biological activities to water or cultured cells electronically, to store such signals on computer disks, and to send them over the Internet (see *Nature* 389, 427; 1997).

A website created by Benveniste's

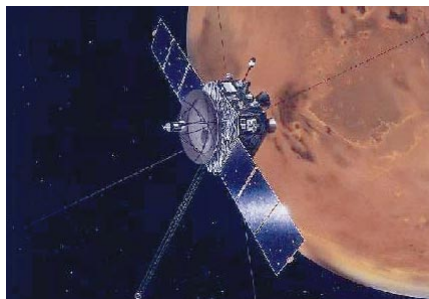
company, Digibio, offers to provide full protocols. "We wish to demonstrate, with the participation of as many laboratories world-wide as possible, the existence of biological activity in a solution so diluted that the number of physical molecules is too low to otherwise induce an effect," says Benveniste (see <http://www.digibio.com>).

British universities back interdisciplinary work

[LONDON] A survey of 2,500 British researchers has found no evidence of systematic bias against interdisciplinary research in the government's Research Assessment Exercise (RAE). The survey, commissioned by the higher education funding councils, concludes that there is therefore no need for major structural changes to the next RAE scheduled for 2001.

Preliminary results from the survey, released at the beginning of the month, show that interdisciplinary research is widespread in British universities, involving three out of four respondents. Eighty-three per cent of heads of departments surveyed felt that overall the RAE has "little effect". But the report concludes that there is a need to strengthen the way that panels assess research quality, particularly in split and 'cross-refereed' departments.

Japan's spacecraft starts long trek to Mars



[TOKYO] Japan's first Mars probe, Nozomi — the Japanese for 'hope' — left the elliptical orbit around Earth 170 days after its launch, to begin its 700 million kilometre journey towards Mars. It is scheduled to enter the Martian atmosphere next October.

Japan's Institute of Space and Astronautical Science announced on Monday (21 December) that Nozomi had been propelled into high orbit. Nozomi will investigate the atmosphere of Mars.

High-tech companies rank skills top priority

[LONDON] Sixty-five per cent of respondents to a survey of British high-technology companies have said that the availability of a

skilled workforce is the most important factor in their decision to locate at a particular place. Proximity to strategic markets, good schools and research facilities were also seen as important.

The survey of 127 life science, telecommunications and information technology companies was conducted by the legal firm Taylor Joynson Garrett and the London Business School. It reveals that more than 40 per cent of small life science and information technology companies have not taken steps to protect the intellectual property rights produced from research.

Canada and Europe broaden research pact

[LONDON] All of the Canadian government's non-defence science and technology activities have been opened up to participation by researchers from member states of the European Union. An agreement signed in Ottawa last week applies to any work funded or performed by government departments or agencies.

In return, Canadian researchers will be able to participate in all areas of the union's fifth Framework programme, which starts next year, although they will have to fund their own work. The deal broadens the scope of an agreement signed in February 1996.

Is Jerusalem the city in Rome's mural?

Sir — Earlier this year, the archaeologist Elisabeta Carnabuci discovered a remarkable painting on the wall of an ancient building in Rome. This mural, which has been described by Nicholas Purcell (*Nature* **392**, 545–547; 1998), may date from about 50 years after the Christian era (AD 50); it shows an aerial view of an unidentified great walled city. There are towers on the walls, and what appears to be a bridge at the upper left of the mural (Fig. 1a). The enclosed city appears to have large open spaces.

But which city can this be? For various reasons, it appears that it is not Rome: for example, Rome did not acquire a wall until 300 years after the Christian era, and, as Purcell notes, the topography of the city in the mural is different from that of Rome — indeed, it is unlike that of any typical Roman city. Purcell has suggested that the city might be one of the other metropolises of the time, such as Alexandria, Antioch or Carthage. I propose that the city might be Jerusalem, on the basis of several similarities between the city in the painting and a map of Jerusalem dating from about the same time (Fig. 1b). If the city really is Jerusalem, then the bridge-like structure is actually an aqueduct (large arrow in Fig. 1b).

Jerusalem, like the city in the mural, is enclosed by a wall adorned by towers (small arrows in Fig. 1b). In particular, notice that both Jerusalem and the unknown city (Fig. 1a) have one tower on one side of the aqueduct or bridge-like structure, respectively, and several towers on the other. Aligning the walls of the two cities, and the aqueduct with the bridge-like structure, it can be seen that the Temple in Jerusalem and the prominent structure on the left of the mural are both at a similar angle.

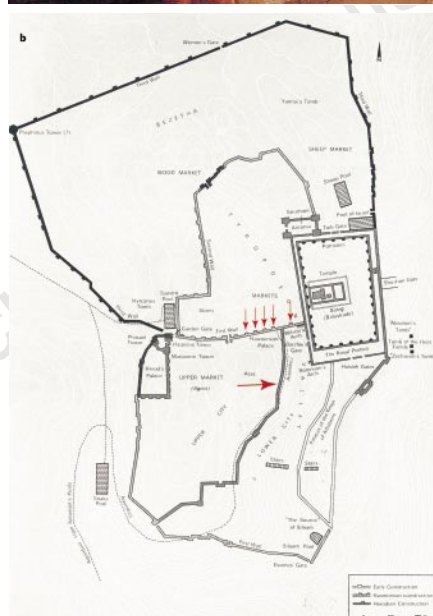
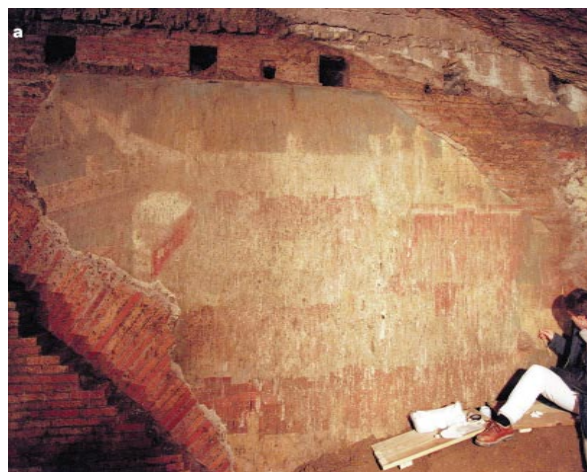
Further excavation of the mural may reveal the true identity of the city, if indeed it does portray a real city. But, of the ancient cities proposed so far, Jerusalem is the best fit. This mural may turn out to be one of the finest depictions of the Holy City.

Eric Lewin Altschuler

School of Medicine, University of California at San Diego, La Jolla, California 92093-0606, USA

Purcell replies — Altschuler's brilliant suggestion has pluses and minuses. One consideration makes it more telling than he knows: the spectacular towers of Jerusalem as rebuilt by King Herod the Great. The historian Josephus described Jerusalem at the time of the appalling siege by the Romans in AD 70, which resulted in its complete destruction, along with Herod's magnificent Temple.

Three enormous towers were key



landmarks, joined by a remarkable network of fortification walls, themselves punctuated by numerous smaller towers, and by the Antonia Fortress. Something of the grandeur of these structures appears from excavations at Herod's palace-fortress at Greater Herodium near Bethlehem. And, in Jerusalem itself, archaeology has revealed soaring terraces, staircases and viaducts where the walls abutted onto the terrace of the Temple. These monuments make a better candidate than Jerusalem's aqueduct for the architecture of the Rome painting.

The epoch of the siege and sacking of Jerusalem fits the date that has been suggested so far for the painting. And few cities would have been so famous in Rome in the last quarter of the first century: the conquest of Judaea was the achievement that legitimated the new dynasty of the emperor Vespasian and his son Titus, and the people of Jerusalem had participated in the colossal triumph that marked the victory. The spoils of the Temple were

Figure 1 Comparison of the Roman wall painting with a map of Jerusalem dating from about the same time. a, The city depicted in the mural. b, A map of Jerusalem before the destruction of the second Temple in AD 70 (map is inverted with respect to a). The small arrows indicate the positions of the towers on the wall; the large arrow shows the aqueduct corresponding to the bridge-like structure in the painting (top left in a). Reproduced from *Encyclopaedia Judaica* (Ketner, Jerusalem, 1996), with permission.

paraded through the streets of Rome, as depicted in the reliefs of the surviving commemorative Arch of Titus. Like the Colosseum, Titus's Baths and the ironically named Temple of Peace, all of which also commemorated the glories of the conquest of Judaea, the arch was in the same part of the city as the painting. The tragedy of the siege, with its terrible suffering, would have been well known.

But there are problems: the building on the left in the painting is certainly a theatre, and, although we know that Herodian Jerusalem had one, it is surprising to see it given so much prominence. The building on the right is a massive temple, and it is just possible to imagine that the configuration of walls, prominent towers, theatre and temple were an attempt at a topographic rendering of Jerusalem at the time of the siege. But — and it's a very big but — the temple doesn't look architecturally much like Herod's Great Temple, about which we are quite well informed. And, above all, it seems to have a huge anthropomorphic statue of a divinity in its courtyard!

So the extent to which the painting is meant to represent Jerusalem, or actually does portray it, is still unclear. Altschuler may be right, and the building on the right is a public building other than the Temple, which was depicted elsewhere on the painting; or the painter may have been extremely ignorant, and only seeking to convey a rough evocation of Jerusalem. Or Altschuler's interpretation may be combined with the view I expressed in my earlier piece: that this is a capriccio, a fantasy city, but one into the creation of which important elements of the wonders of Herodian Jerusalem have been incorporated.

Nicholas Purcell

St John's College, Oxford OX1 3JP, UK

Longevity and the barren aristocrat

Daniel E. L. Promislow

Experiments with fruitflies have shown that those with fewer offspring generally live longer. Twelve-hundred years of genealogical data from British aristocrats now suggests that this relationship extends to humans.

When Madame Jeanne Calment (Fig. 1) died last year, at a record-breaking 122 years of age, gerontologists debated just what accounted for her extraordinary longevity. According to a report by Westendorp and Kirkwood¹ on page 743 of this issue, at least part of Madame Calment's Methuselian lifespan may have been due to her having had only one child. These authors used data from genealogies of the British aristocracy to show that lifespan is negatively correlated with family size — a result that supports predictions from evolutionary models of ageing.

Ageing is an inevitable result of the declining force of natural selection with age². For example, a mutant gene that halves the survival rate among young children will be strongly selected against. But a gene of similar magnitude whose effects are confined to people over the age of 90 will experience no selection because people with this deleterious gene will have passed it to their offspring long before its effects become apparent. So, over evolutionary time, late-acting deleterious genes will accumulate, leading to an increase in mortality rates late in life. Indeed, late-acting deleterious genes may even be favoured by selection if they have beneficial effects early in life — a genetic mechanism dubbed 'antagonistic pleiotropy'³. This trade-off between benefits at an early age and costs later was explained by Kirkwood over

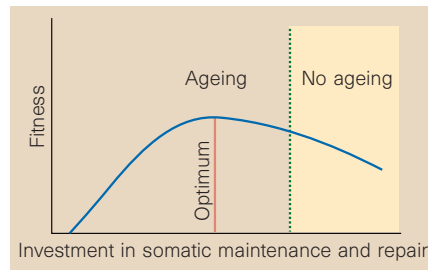


Figure 2 Darwinian fitness increases with increasing rates of reproduction and survival. According to Kirkwood's 'disposable soma' theory⁴, as investment in cell maintenance and repair increases, survival rates increase but reproductive rates decline. With enough investment in maintenance and repair, no ageing will occur. However, Darwinian fitness is maximized at some intermediate optimum, where cellular investment is enough to allow for survival to reproduction, but insufficient to eliminate ageing. (Figure adapted from ref. 11.)

20 years ago in his 'disposable soma' theory⁴, which predicts that any investment in reproduction diverts resources away from the maintenance and repair of cells, with ageing as a result (Fig. 2).

The antagonistic-pleiotropy and disposable-soma theories both indicate that those who invest heavily in reproduction while young will pay for this reproductive success with a shortened lifespan. Researchers have

found support for this prediction in many model organisms. In the fruitfly *Drosophila melanogaster*, for example, flies selected for long lifespan have reduced early fecundity⁵.

Now, thanks to the extensive genealogical records that have been kept on the British aristocracy for the past 1,200 years, Westendorp and Kirkwood¹ have identified the predicted trade-off in humans. The authors obtained records for 19,830 male and 13,667 female aristocrats born between 740 and 1875, then looked at the dates of birth and death, number of children, and the age at which each child was born. For women who survived to at least 60 years of age, they found a negative correlation between the number of years survived beyond 60 and the number of children. (The pattern among men was virtually identical.) Similarly, women who started reproducing later (and, presumably, had smaller families) lived longer. Almost half the women who surpassed 81 years of age had no children, yet fewer than one-third of the women that died before that age were childless.

One might argue that the negative correlation between family size and longevity is due to changing demographic trends — starting around 1700, average family size decreased and longevity increased among the British aristocracy. To ensure that this demographic transition was not the underlying cause of the negative correlation between longevity and the number of children, the authors analysed cohorts born before 1700 and after this time separately. In both cases, the correlation between lifespan and number of progeny was negative, although evidence for the trade-off was much greater among women born before 1700 than those born more recently.

This study, and previous tests of evolutionary theories for ageing using human data^{6,7}, assumes that the observed variation in longevity is due, at least in part, to genetics. Westendorp and Kirkwood show that longevity is correlated between parents and offspring, suggesting a heritable basis to ageing. However, for genealogical studies at least, we cannot distinguish the effects of genetic similarity from those of common environment. The authors find that longevity is correlated not only between parents and offspring, but also between husband and wife. So, inbreeding aside, it is fair to assume that any correlation of longevity between spouses is due to a shared environment, not shared genes.

Westendorp and Kirkwood reduce potential environmental variation by limiting their study to the aristocracy. But even among this relatively homogenous class one might expect substantial variation in wealth and general environment — not only across the centuries, but even among families living in the same period. The new results are powerful evidence for a correlation between



Figure 1 Madame Jeanne Calment, on her 121st birthday, and her identity card.

reproduction and longevity in British aristocrats. In the search for trade-offs, however, correlation does not prove causation⁸, and it remains to be seen whether the observed patterns are due to nature or nurture.

Nevertheless, Westendorp and Kirkwood's results raise intriguing questions that might be addressed with this extraordinary human database. First, for half a century, evolutionary biologists studying life-history strategies have been interested in the evolution of clutch size (number of progeny)⁹. The database can now be used to determine whether there is an optimal clutch size (such as the number of grand-offspring) that maximizes Darwinian fitness in humans. Second, at least in the twentieth century, married men and women tend to outlive their single counterparts. Although the authors confined their analysis to married adults, it would be interesting to determine the effect of marital status on longevity over the centuries. Third, this database may also be useful for those interested in the causes and consequences of variation in female reproductive behaviour. For example, one surprising result from Westendorp and Kirkwood's study was that almost half of the women who survived to age 81 or older had no children at all. What could be the biological and socioeconomic explanations for why so many married women remained childless?

Can we conclude that having fewer children extends life expectancy? Although childless women benefit from a decreased incidence of heart disease and cervical cancer later in life, they have an increased risk of breast cancer and respiratory disease¹⁰. Clearly, many factors influence lifespan, and readers can rest assured that having more children will not necessarily send you to an early grave. After all, Queen Victoria gave birth to nine offspring, all of whom survived, and lived to be 81 years old — in her day, a ripe old age. □

Daniel E. L. Promislow is in the Department of Genetics, University of Georgia, Athens, Georgia 30602-7223, USA.

e-mail: promislow@bscr.uga.edu

1. Westendorp, R. G. J. & Kirkwood, T. B. L. *Nature* **396**, 743–746 (1998).
2. Medawar, P. B. *An Unsolved Problem in Biology* (H. K. Lewis, London, 1952).
3. Williams, G. C. *Evolution* **11**, 398–411 (1957).
4. Kirkwood, T. B. L. *Nature* **270**, 301–304 (1977).
5. Rose, M. R. & Charlesworth, B. *Nature* **287**, 141–142 (1980).
6. Gavrilova, N. S. et al. *Hum. Biol.* **70**, 799–804 (1998).
7. Le Bourg, E., Thon, B., Legare, J., Desjardins, B. & Charbonneau, H. *Exp. Gerontol.* **28**, 217–232 (1993).
8. Reznick, D. *Oikos* **44**, 257–267 (1985).
9. Lack, D. *The Natural Regulation of Animal Numbers* (Clarendon, Oxford, 1954).
10. Kvale, G., Heuch, I. & Nilssen, S. *Int. J. Epidemiol.* **23**, 691–699 (1994).
11. Kirkwood, T. B. L. & Cremer, T. *Hum. Genet.* **60**, 101–121 (1982).

Oceanography

Something stirs in the deep

Peter Killworth

In a classic paper entitled “Abyssal recipes”, published in 1966, Walter Munk¹ tackled one of physical oceanography's big questions. How do the cold oceanic waters that sink into the abyss at high latitudes, and flow into low latitudes, ever rise again as warm water? The simple answer is that the deep waters are subject to mixing, but putting mathematical clothes on that answer is immensely tricky. Together with Carl Wunsch², Munk has now returned to the issue in “Abyssal Recipes II”. The new answers are again highly provisional, but they provide a provocative agenda for climatologists as well as oceanographers.

The top few metres of the ocean have the same heat capacity as the entire atmosphere, so that the upper ocean can easily both absorb and drive atmospheric heat changes. It is hardly surprising, then, that climate researchers care about what happens at the ocean surface. But the deep ocean would seem to be different. It is far away from contact with the atmosphere both in space and in time; a water parcel starting at 5 km depth would take a century to rise through the ocean vertically. Yet the deep ocean has a strong influence on world climate. In a few

thousand years, the dense, cold, salty water that forms in high latitudes, and sinks to great depths, ought to fill up all the ocean basins from underneath. The result would be

an ocean comprised almost completely of such water, which would have a dramatic effect on the climate.

This doesn't happen, because the deep ocean is stirred. But how? In “Abyssal recipes” Munk sought a balance between the upwelling of cold water with mixing down of warm water by processes that could be represented by some turbulent diffusion mechanism. Munk found that a canonical value of $1 \text{ cm}^2 \text{ s}^{-1}$ for the diffusion coefficient, equivalent to a vertical spreading from a point source of 50 m in one year, would account for heat and salt observations in the deep ocean. The method he used was later shown to be numerically ill-posed³, but the diffusivity value has nonetheless endured.

In “Abyssal recipes II”, Munk and Wunsch² come armed with a barrage of new methods and observations. The predominant water-exchange balance between cold upwelling and warm downwards mixing remains the same. We no longer believe the ocean to be upwelling uniformly, however, and the ‘diffusion’ may well be a surrogate for many localized active regions such as well-stirred bottom slopes. Certainly the experimental evidence from measurements of oceanic microstructure⁴ and dye releases⁵ is that, over much of the deep ocean, diffusivities are an order of magnitude smaller than Munk's initial estimates. Munk and Wunsch estimate the energy required to maintain the low (large-scale) and high (localized) mixing at 2.1 terawatts (TW). Two questions are thus raised: how is this energy generated, and what oceanic processes would actually cause the mixing to occur?

At first sight, there are many obvious candidates for the energy source: solar radiation is 100,000 times greater, and poleward heat flow in the ocean or atmosphere a thousand times greater, than needed. Curiously,

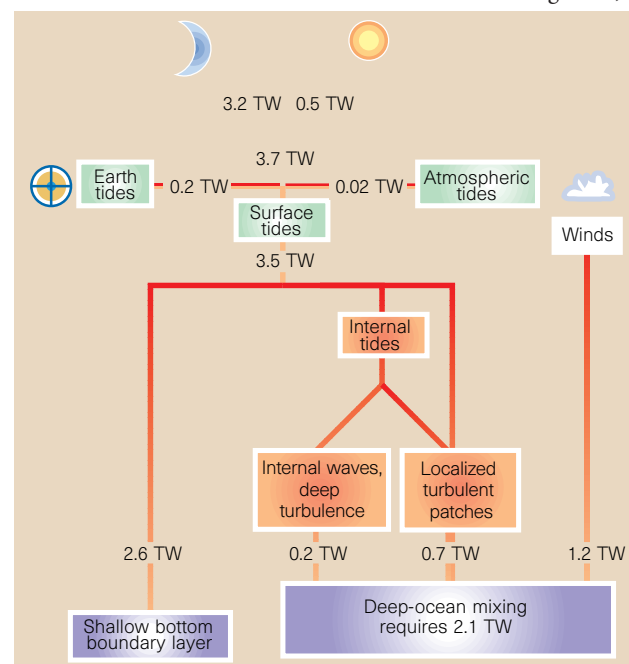


Figure 1 Lunatic estimates? This is a simplified form of the tidal energy flux budget that Munk and Wunsch² estimate delivers 2.1 terawatts (TW) of energy — largely from the Moon and surface winds but with a small component from the Sun — for deep-ocean mixing. There are however many uncertainties in the budget, especially in the estimates of the energies of the oceans' internal tides. (Modified from ref. 2.)

although the ocean is indeed partly driven by such buoyancy forces (of order 1 TW), they cannot drive the mixing because the ocean is an inefficient heat engine; that is, differences in temperature are not transformed effectively into work done in the form of moving water masses.

Instead, say Munk and Wunsch, the prime candidates appear to be more mechanical: tidal dissipation, largely powered by the Moon, and wind driving (Fig. 1). Tidal dissipation has usually been assumed to occur predominantly in bottom boundary layers on shallow ocean shelves, because dissipation varies as the cube of the water speed, and ocean tidal velocities tend to vary inversely with ocean depth. Enough dissipation appears to be left over, however, in the form of internal tides (about which little is known) to provide about 1 TW of power, some in the form of turbulence away from the ocean floor, but most as turbulent patches through scattering by topography. The remainder of the energy is input directly as work done by wind at the ocean surface (another 1.2 TW; see Fig. 1).

So, although with many uncertainties, energy sources for deep-ocean mixing are fairly clear. The variety of mechanisms involved are much less so. Boundary mixing — preferentially high mixing due to turbulence in bottom and side boundary layers — has long been a candidate, and theories relating such mixing to effective internal ocean diffusivities exist. But there are few direct measurements of these processes and caution is necessary in extrapolating from what little we do know. In a manner that we do not understand, there are regions in which

inferred diffusivities are several orders of magnitude higher than background⁶, and they are usually located over steep structures on the ocean floor. This is in line with beliefs that tidal energy spreads from ridges and other topographic features — and there is a lot of topography in the ocean (over half a million seamounts in the Pacific alone).

Almost everything in “Abyssal recipes II” remains uncertain: oceanographers have no way to survey large tracts of the ocean interior; the possible range of both processes and locations is huge; and estimates of the power generation necessary have wide error bars. We need modelling work to see how the processes can add up. Such modelling has to be both process-oriented (to explore the various possibilities) and heavily numerical, including some or all of the processes — for example, no general circulation model I am aware of contains tidal mixing.

To cap it all nicely, Munk and Wunsch² note in passing: “To many readers, the proposal that the Moon plays a major role in the general [ocean] circulation will border on the lunatic”. Maybe, maybe not. It’s a lovely concept. □

Peter Killworth is at the Southampton Oceanography Centre, Empress Dock, Southampton SO14 3ZH, UK.

e-mail: P.Killworth@soc.soton.ac.uk

1. Munk, W. *Deep-Sea Res.* **13**, 707–730 (1966).
2. Munk, W. & Wunsch, C. *Deep-Sea Res.* **45**, 1976–2009 (1998).
3. Olbers, D. J. & Wenzel, M. in *Oceanic Circulation Models: Combining Data and Dynamics* (eds Anderson, D. L. T. & Willebrand, J.), 95–139 (Kluwer, Dordrecht, 1989).
4. Gregg, M. C. *J. Geophys. Res.* **94**, 9686–9698 (1989).
5. Ledwell, J. R., Watson, A. J. & Law, C. S. *Nature* **364**, 701–703 (1993).
6. Polzin, K. L., Toole, J. M., Ledwell, J. R. & Schmitt, R. W. *Science* **276**, 93–96 (1997).

Auditory perception

Sounds in a virtual world

Malcolm N. Semple

Sounds delivered through headphones can be made to seem as though they originate from sources outside the head — a compelling illusion known as virtual auditory space. Although current implementations are far from perfect, progress has been encouraging¹ and practical applications could include enhancing the information content of virtual environments such as military simulations or computer games. Virtual auditory space also allows auditory researchers to control stimuli very precisely and to investigate stimuli, such as moving sounds, that are difficult to generate using free-field speakers.

On page 747 of this issue, Kulkarni and Colburn² provide new insight into the factors that underlie the externalized perception of a sound source. In principle, for a listener to perceive virtual auditory space as real, the normal cues for directional hearing

must be simulated accurately through headphones. What are the cues? There has been a long history of investigation into how we detect where a sound is coming from³, dating back to the work of Lord Rayleigh in the last century. Psychoacoustic and physiological studies emphasized two main cues for detecting the direction of a sound source with both ears: the time delay between the sound reaching one ear and the other; and the difference in sound level at the two ears (Fig. 1, overleaf). These ‘binaural difference cues’ are computed within the brain after the information from each ear converges there.

Over the past decade there has been a growing awareness that the characteristics of the external ears also contribute to our perception of where a sound is coming from. ‘Spectral cues’ are generated because the outer part of the ear, the cone-shaped pinna, applies a filter function — the ‘head-related



100 YEARS AGO

The proposed application of electrical power for mounting plays at Drury Lane, on the lines advocated by Mr. Edwin O. Sachs, has now taken a tangible form in the completion of the first section of the stage installation in time for the impending pantomime. Mr Sachs’ present work refers principally to the stage floor and its movability in sections above and below the footlights. The total area now already movable by mechanical power exceeds 1200 square feet. The electrical appliances just completed take the form of so-called “bridges,” each working independently. ... They can travel about 20 feet vertically.

From *Nature* 29 December 1898.

50 YEARS AGO

J. Bardeen and W. H. Brattain, of the Bell Telephone Laboratories, in the course of general investigation, initiated and directed by W. H. Shockley, on the properties of semiconductors, have developed a new three-element electronic device, called by them a ‘transistor’ (transfer-resistor) or ‘semi-conductor triode’. This can be employed as an amplifier or oscillator and can replace the vacuum triode in most electronic circuits. The device, which consists of a metal cylinder, about an inch long and of the thickness of a pencil, containing the germanium and its three electrodes, was demonstrated with great success recently at a Press conference in New York. ... In place of the single ‘cat’s whisker’ of the crystal rectifier, two fine wires (both tungsten and phosphor-bronze have been used) make contact with the upper surface of the germanium block. These electrodes, the emitter and collector, are about 0.005–0.025 cm apart. The third electrode, which is a large-area low resistance contact, makes contact with the base of the block. ... Amplifications up to 100 times, that is to say, a gain of 20 decibels, have been obtained. The transistor uses less power than a vacuum tube, has an output of 25 milliwatts and can operate at frequencies up to 10 megacycles per second. ... The transistor is not yet in production, but its simplicity, small size, performance, long life, and probably low cost when mass-produced, should find many applications in all forms of electronic equipment.

From *Nature* 25 December 1948.

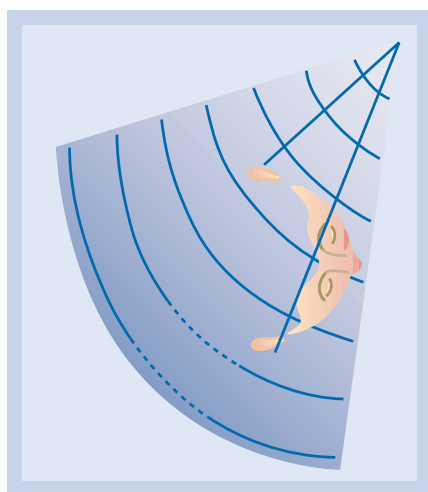


Figure 1 The two ears, like the eyes, capture the world from slightly different perspectives. When a sound is displaced to one side of the head it arrives earlier at the near ear, leading to a time difference between the two ears. If the wavelength is short relative to the size of the head, reflection and refraction will attenuate the signal en route to the far ear, leading to a difference in the level between the ears.

transfer function' (HRTF) — to the sound, shaping its amplitude and phase spectra (Fig. 2). There is now a general consensus that these external-ear spectral cues provide information about the location of a sound source in the vertical plane (or 'elevation')⁴, and that the binaural difference cues provide a basis for discerning the horizontal angle (or 'azimuth') of the sound. Although the effects of the external ear are commonly described as being 'monaural spectral cues',

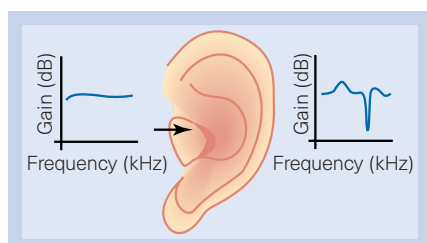


Figure 2 Filtering of sound by the external ear. As sound is funnelled to the ear canal, it is subjected to a frequency-dependent filtering that modifies the phase and amplitude spectra. This filtering is known as the 'head-related transfer function' (HRTF), and Kulkarni and Colburn² measured it by inserting a microphone probe tube deep into the ear canal, to sample signals near the eardrum. They then compared these signals with signals recorded near the entrance to the ear. The measurement was repeated for sounds coming from a range of locations.

the cue is not exclusively from one ear because the spectral notches shape each ear's input to the neurons that process signals from both ears.

The impression of a sound with direction can be generated through stereophonic headphones by incorporating differences in the timing and intensity of the signals at the two ears. But these auditory images are typically confined within the head. Externalized images can be elicited only by carefully compensating for any spectral coloration generated by the headphones, and by applying a filter function to compensate for the normal HRTF. Hartmann and Wittenberg⁵ were the first to investigate the factors that are important for externalizing the acoustic image in a

controlled way, and they found that externalization requires realistic spectral profiles in each ear — that is, the sound experienced at the eardrum should reproduce the sound that would occur naturally. Because there is considerable variability in the size and shape of the human pinna, such a realistic virtual space image requires that the listener's own HRTF is measured⁶. In fact, when people try to listen via the filtering of another person's ears, their ability to distinguish where a sound is coming from suffers markedly⁷. Moreover, there is evidence that in the normal course of repositioning the headphones, the HRTF may be shifted considerably.

These are serious obstacles to the widespread use of virtual auditory space, because measuring the HRTF is time-consuming and tiring. One hope is that by learning more about which details of the HRTF are responsible for externalizing the acoustic image, we may be able to avoid having to include every last detail of the HRTF in the calculations. Kulkarni and Colburn² have now made considerable progress towards this goal. Using an open-earphone technique that allows sound to be presented either from a real free-field source or as a virtual image, the authors asked listeners to discriminate between real and virtual sounds. Usually, virtual sounds are given through headphones, which must then be removed for presentation of real sounds. Instead, Kulkarni and Colburn delivered the virtual sounds into the ear canal through small tubes so that, without removing the tube-phones, sounds originating from an external speaker could enter the ear unimpeded. They could then immediately compare how accurately a virtual sound reproduced the real sound that it was designed to simulate.

The authors found that listeners cannot discriminate real from virtual sounds presented in this way. Moreover, the virtual sound remained perceptually identical to the real sound — even after most of the fine detail of the HRTF (both phase and magnitude) was removed. This does not mean that an individual HRTF is unimportant. But, for an accurate reproduction of real three-dimensional auditory space, it seems that a virtual space signal must compensate for only the most prominent features of the individual listener's HRTF.

Earlier this year it was shown⁸ that if the pinna is modified using ear moulds, the ability to localize sounds (particularly in elevation) is dramatically reduced. But people can re-learn localization if they are given time to adjust to these new ears. Moreover, they can still localize sounds accurately using their own ears, immediately after the moulds have been removed. Perhaps there is a use for virtual environments that distort, rather than faithfully replicate, real auditory space. For example, by exaggerating the magnitude of the cues, it might be possible to train an

Electronic databases

It's good to talk

These days, electronic databases are the quickest, easiest way to search the scientific literature. But how good are they at finding all of the relevant studies? According to a study by F. D. R. Hobbs and colleagues (*Br. Med. J.* 317, 1562–1563; 1998), in developing fields at least, it may be better to rely on word of mouth.

The best way to find all studies relevant to a particular field is to hand-search the literature. In most fields this is difficult enough, but imagine the task in new fields, where the specialist literature has not become established and papers are widely scattered.

This is the challenge that Hobbs and colleagues took on. Having chosen a relatively new field known as 'near patient testing', they sent questionnaires to academics and commercial companies, requesting key references from journals and the names of other workers in the field. They then compared the answers



with the results of searching eight databases.

Their results confirmed previous studies which showed that electronic databases may uncover only half of all relevant studies. Indeed, almost a quarter of the 'eligible' references would not have been discovered without the help of experts in the field. It seems that, in developing fields, there's still no substitute for seeking expert advice. **Alison Mitchell**

observer to use a broader range of spatial channels.

The recent progress made in this field is not limited to human studies — some groups are using virtual auditory space methods to study the brain mechanisms for sound localization in anaesthetized animals⁹. By exploring neural responses to virtual-space signals, complex properties of receptive fields are revealed without the limitations inherent in using an array of speakers or a mobile sound source. However, because we have no behavioural measure of the animal's perception of the virtual sound, and because we cannot present both the virtual signal and its real counterpart, we do not know how well the virtual stimulus replicates real auditory space. These concerns could be alleviated by adapting the approach of

Kulkarni and Colburn to physiological studies in behaving animals. □

Malcolm N. Semple is at the Center for Neural Science, New York University, 6 Washington Place, New York, New York 10003, USA.
e-mail: mal@cns.nyu.edu

1. Carlile, S. *Virtual Auditory Space: Generation and Applications* (Chapman & Hall, New York, 1996).
2. Kulkarni, A. & Colburn, H. S. *Nature* **396**, 747–749 (1998).
3. Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, Massachusetts, 1983).
4. Gardner, M. B. & Gardner, R. S. *J. Acoust. Soc. Am.* **53**, 400–408 (1973).
5. Hartmann, W. M. & Wittenberg, A. *J. Acoust. Soc. Am.* **99**, 3678–3688 (1996).
6. Wightman, F. L. & Kistler, D. J. *Nature Neurosci.* **1**, 337–339 (1998).
7. Wenzel, E. M., Arruda, M., Kistler, D. J. & Wightman, F. L. *J. Acoust. Soc. Am.* **94**, 111–123 (1993).
8. Hofman, P. M., Van Riswick, J. G. A. & Van Opstal, A. J. *Nature Neurosci.* **1**, 417–421 (1998).
9. Brugge, J. F. *et al. Hearing Res.* **73**, 67–84 (1994).

Nuclear fusion

Advanced fuels under debate

G. L. Kulcinski and J. F. Santarius

The energy released by the fusion of deuterium and tritium has long been touted as a replacement for power plants that burn fossil fuels (which, of course, emit huge amounts of greenhouse gases), and for fission reactors which generate long-lived, high-level radioactive waste. Not only is fusion technology struggling to overcome the difficulty in confining a deuterium (D) and tritium (T) plasma, and to attain commercial viability, it is struggling with the consequences of intense neutron bombardment of confining structures. Fuels that release far fewer neutrons would offer considerable advantages over the D–T cycle — hence a meeting* on the topic last month.

Deuterium and tritium are the principal first-generation fusion fuels, because they burn most easily. Alternative fuel cycles are being given serious thought because of the radiation damage and radioactive waste products that result from the D–T cycle, which emits 80% of its energy in high-energy neutrons. These neutrons are known to cause severe degradation of the structural components and generate large quantities of radioisotopes; D–T fusion produces four times as many neutrons as fission reactors per kilowatt hour of thermal energy released.

Advanced fusion fuels are defined as those whose neutron production rates are very low, or even zero. The second-generation fusion fuel cycles (for example deuterium–helium-3; see Fig. 1) release only a few percent of their energy in neutrons, whereas the third-generation fuels (proton–boron-11, helium-3–helium-3 and others) are

essentially neutron free. In magnetic fusion technology, confinement of the fuel in the form of a plasma (ionized gas) is a central issue. The increased profile of advanced fuels coincides with the re-emergence of innovative confinement concepts, which are particularly suited to burn such fuels.

One of the main advantages of the second- and third-generation fuels is that they greatly reduce the radiation damage to fusion chamber structures, allowing these components to last the full lifetime of the power plant. These fuel cycles produce little or no long-lived radioactivity, thus reducing the expense of decommissioning a plant when its working life is over. They also release a large fraction of their energy in the form of charged particles, so they would allow direct conversion of fusion-product energy to electricity at efficiencies of 70% or higher. This is roughly twice what the first-generation fuels will attain with thermal conversion — the use of neutron-heated coolants to create superheated steam to drive

turbines.

The main disadvantage of the second-generation fuels is that they require 4–5 times higher temperatures and 4–5 times better confinement conditions to compete with the first-generation fuels. The third-generation fuels require factors of more than 20. Generally, the second- and third-generation fuel cycles have lower fusion power densities. That is the reason why advanced fuels have to be coupled with new confinement techniques, such as those discussed at the meeting. Such techniques could have an intrinsically higher power density than the tokamak, currently the mainline approach pursued around the world.

One issue raised was that engineers might not be able to take full advantage of D–T fuel in the new, high-power-density approaches because of high neutron wall loads. The ³He fuel cycles also have the problem of the location of a long-lasting fuel source. Although there is enough ³He in the United States and Russia to conduct all the research needed to develop the first commercial fusion power plant, that amounts only to some hundreds of kilograms. Subsequently, much greater quantities (tens of tonnes per year) would be required for a worldwide fusion reactor economy, but the main relevant source (at around 1,000,000 tonnes) is on the Moon.

Four concepts especially suited to burning the second- and, in some cases, the third-generation fuels were aired at the meeting. A crucial distinction here is between Maxwellian and non-Maxwellian plasmas. Collisions force plasmas nearly to thermodynamic equilibrium, except for usually small flows through the physical boundaries, and the resulting plasma is said to have a Maxwellian distribution. Some concepts rely on input power or pulsed operation to keep the system far from equilibrium, and these plasmas are called non-Maxwellian.

The first of the approaches discussed has been proposed in a near-term, high-field, D–³He experiment called CANDOR, based on Maxwellian plasma in a tokamak (B. Coppi and L. E. Sugiyama, Massachusetts

First generation	D + T	→	n + ⁴ He	17.6 MeV
	D + D	→	n + ³ He	3.65 MeV (av.)
		→	p + T	
Second generation	D + ³ He	→	p + ⁴ He	18.4 MeV
Third generation	p + ¹¹ B	→	3 ⁴ He	8.7 MeV
	³ He + ³ He	→	2p + ⁴ He	12.9 MeV

Figure 1 The main fusion reactions, and their energy yield. For instance, fusion of deuterium (D) and tritium (T) yields a neutron and helium-4, with release of 17.6 mega electronvolts of energy. The second- and third-generation fuel cycles result in very small or no release of neutrons (neutrons damage plasma-confinement structures — hence the reason, among others, for investigating the alternative fuels).

*Advanced Fuels for Fusion, miniconference at the Annual Meeting of the American Physical Society's Division of Plasma Physics (http://www.aps.org/meet/DPP98/), New Orleans, 19 November 1998.



Figure 2 Plasma in an inertial-electrostatic confinement device at the University of Wisconsin. This type of device has provided proof-in-principle that fusion can be produced with $D-^3He$, a second-generation fuel, at low levels of power input.

Institute of Technology). A D-T version, IGNITOR, is under scrutiny as a possible option for the next step in the US fusion programme. Another Maxwellian $D-^3He$ plasma concept is based on confinement using a dipole magnetic field produced by a single superconducting ring magnet (L. Bromberg, MIT); a modest-sized dipole experiment aimed at exploring it further has recently begun construction at MIT.

These two lines of thinking illustrate the main routes to advanced-fuel fusion using Maxwellian plasmas: high magnetic field or high values of beta, the ratio of plasma pressure to magnetic-field pressure. Several high-beta confinement concepts were mentioned briefly at the miniconference, such as the field-reversed configuration, spheromak, spherical torus and reversed-field pinch.

The third approach to be debated was inertial-electrostatic confinement (Fig. 2), which has both non-Maxwellian and oscillating Maxwellian embodiments (D. Barnes, Los Alamos National Lab.). In these devices, for which exploratory experiments have been successful, electrons in a Penning trap form spherically-symmetric electrostatic potential wells, producing a radially convergent ion flow. The very high plasma density that results may even allow burning of a D-D fuel. Finally, there is the idea of a colliding-beam reactor, which is based on maintaining large-orbit ion currents circulating in opposite directions (N. Rostoker, Univ. California, Irvine). This method might be used in either a pure colliding-beam form or to induce stability in field-reversed configurations.

The fireworks in the ensuing discussions showed that there is a sharp division of opinion about such basic principles as net energy production and engineering feasibility, and thus about the prospects for the new fuels

and confinement techniques. Some arguments centred around the technology needed to drive the plasmas (M. Lampf and W. M. Manheimer, Naval Research Laboratory); others on the accuracy of the calculations of net energy production in both Maxwellian and non-Maxwellian plasmas (T. H. Rider, MIT; W. M. Nevins, Lawrence Livermore National Lab.). There was general agreement that $D-^3He$ (but probably not $p-^{11}B$) fuel could be burned in Maxwellian plasmas. For non-Maxwellian plasmas, however, there was a disagreement between proponents and opponents of the approach. Several of the proponents discussed power plants that would rely on non-Maxwellian plasmas where the calculations of collisional effects, and of the energy losses due to the emission of *bremsstrahlung* and synchrotron radiation, are difficult and subject to debate.

Neither side yielded on the prospects for the second- and third-generation fuels. But there was general consensus on the overwhelming benefits if they could be made to operate efficiently. Even the task of procuring 3He from the Moon did not raise eyebrows, perhaps because many NASA reviews have assessed the engineering approaches to such an undertaking and found them to be feasible.

The session ended on the same question on which it began — can at least one of the alternative concepts demonstrate net energy production with an advanced fuel? The answer seems to be a long way off at the present level of theoretical and experimental research. At the least, however, the two sides are now confronting the issues. □

G. L. Kulcinski and J. F. Santarius are in the Fusion Technology Institute, University of Wisconsin-Madison, 1500 Engineering Drive, Madison, Wisconsin 53706, USA.
e-mail: kulcinski@engr.wisc.edu

Daedalus

Brandy and cigars

Smoking has gone from the height of chic to the depth of infamy in only a few decades. And yet, compared to other addictive drugs, tobacco is socially almost harmless. Its users do not become career criminals, loud-mouthed tearaways or menaces on the roads. They damage only themselves; their only anti-social behaviour is to spread the nasty smell of smoke.

And even that nasty smell, says Daedalus, is not central to the habit. It just so happens that free nicotine, the addictive agent in tobacco, reacts readily with oxygen. Tobacco itself contains nicotine in combination, largely as the stable citrate. Smoking is a way of freeing it on demand, and getting it into the smoker in seconds, before it has a chance to oxidize. Daedalus is therefore devising an oxygen-free vehicle for nicotine.

His brilliant idea is to combine it with another widely used drug, alcohol. Brewing, of course, exploits the conversion of sugars to alcohol. The reaction needs no oxygen; indeed, it gives off carbon dioxide which tends to keep air away. Many sources of plant sugar, from potatoes to barley to grapes to elderflowers, can be fermented to an alcoholic wine or beer. So DREADCO chemists are now inventing tobacco wine. Their crucial new technique is to conduct the process from the very beginning in an atmosphere of carbon dioxide. No oxygen can get in, even in the early stages. Once the product is bottled, of course, the oxygen problem is essentially solved. Tiny traces of the gas may slowly diffuse in through the cork; but as in conventional wines this will probably go to producing higher esters and aldehydes which enhance the flavour and bouquet.

Tobacco wine should be widely welcomed. Unlike the various distressing 'improved cigarettes' devised by the tobacco companies over the years, it will fill a well-understood niche in the market. It will probably not be very alcoholic. With a bottle or can of tobacco wine at their side, smokers will find it easy to give up their cigarettes. They will be able to enjoy their addiction without the tars and combustion-products which make smoking hazardous and unpopular. Indeed, they will be able to deny and sublimate their addiction in the manner perfected by innumerable alcoholics. They will develop an exaggerated, connoisseur's awareness and appreciation of infinitesimal subtleties of smell and flavour in the product.

David Jones

Frémiet's frenzy

Nineteenth-century museum sculptures depicting great moments in prehistory might seem kitsch to modern eyes. But, in their time, they were a key way for the latest scientific ideas to be communicated to the public.

Martin Kemp

The Jardin des Plantes in Paris is far more than a garden of plants. In the heyday of Lamarck, Geoffroy de Saint-Hilaire and Cuvier, it was an international crucible for scientific natural history, and the centre for controversies about the great prehistoric transformations that studies of the fossil record were revealing. Comparative anatomy, the chair occupied by Cuvier from 1795, was at the centre of the intellectual conflagrations.

In the Jardin today, the Galerie d'Anatomie Comparée confronts the visitor with a striking testimony to Cuvier's legacy and to its subsequent Darwinian transformation. Constructed by the architect Ferdinand Dutert, the gallery opened in 1898 with displays devised by Albert Gaudry, the leading French advocate of Darwin's theories. On the first floor are the palaeontological collections, arranged according to evolutionary principles, while on the ground floor the visitor is immediately faced by a veritable Noah's Ark of marching skeletons, headed triumphantly by a painted *écorché* statue (an anatomical figure in which the muscles are revealed) of an upwardly aspiring man.

Less regarded today, but no less integral to the gallery's message, are the sculptural adornments found outside and inside the building. The leading participant was



Follow me: *Ecorché Man Leading the Parade of Animal Skeletons*, from the Galerie d'Anatomie Comparée, Jardin des Plantes, Paris.



Emmanuel Frémiet, *An Orang-utan Strangling a Native of Borneo*, 1898, in the foyer of the Galerie d'Anatomie Comparée.

the specialist sculptor of military and animal subjects, Emmanuel Frémiet, who had succeeded the great *animalier*, Antoine-Louis Barye, as Professeur d'Iconographie Naturelle at the Jardin in 1875.

Famed for his equestrian monument to Joan of Arc in the Place des Pyramides, Frémiet's animal tableaux courted controversy. His favourite subjects represented humans and animals in frenzied competition: initially as realizations of French theories of the dynamic march of human history, latterly in a Darwinian context.

In the foyer is one of his stock images of natural violence, *An Orang-utan Strangling a Native of Borneo*, originally enhanced with polychrome touches of gore. One of his exterior bronze reliefs, on the facade overlooking the Rue Buffon, depicts a key combat in evolutionary iconography, the hard-won triumph of man over bear. The relief itself, scoured by acid rain, can now best be appreciated via the smaller version adorning the reverse of the plinth of the memorial statue erected in front of the gallery.

Victory over the bear, not least in the battle for supremacy as cave-dwellers, was seen as a vital step in human progress. Louis Figuier's popular *L'homme primitif* in its 1870 edition illustrates competition with the 'great bear' which lends its name to the first of his stone-age epochs. It was the advent of tools, such as the hafted axe slung over the shoulder of Frémiet's Herculean caveman, that enabled humans to "repulse the attacks of ferocious animals which prowled around his retreat and often assailed him," to quote the English translation of 1872.

In the smaller bronze, one of the sculp-



Emmanuel Frémiet, *Man Triumphant over Two Bears*, Jardin des Plantes.

tor's bears hangs limply over the base of the relief, mortally wounded by a broken spear, while a youngster is hauled away, literally, by the scruff of its neck. The heroic man strides forth, beard thrust resolutely into the distance, marching irresistibly towards the rising sun of the new age (a detail less apparent in the larger relief).

We may now be inclined to see such images as pieces of decorative nonsense, if we notice them at all. In fact, they were as important in the public understanding of science as any of the museum's displays, and have, particularly through related book illustrations, done much to fire the public imagination and inspire fledgling palaeontologists. We exclude such works from our concept of mainstream scientific endeavour at the risk of severely misunderstanding the public dynamic of the scientific vision. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Spines and tissues of ancient sharks

The 'spine-brush complex'¹ of the extinct, mid-Palaeozoic primitive chondrichthyan shark *Stethacanthus* is one of the strangest vertebrate appendages known. Its structure has never been defined, but here we reveal that the 'brush' is actually an enlarged, specialized extension of the fin baseplate (basal cartilage²). It consists of an

unusual type of globular calcified cartilage, a tissue that is often associated with pre-jawed primitive vertebrates³.

The brush (Fig. 1a) is an apparently fibrous, fan-shaped support for a platform of large, tooth-like scales. The preceding spine is bony and rides on a large baseplate (Fig. 1a). The spine, like those of modern

chimaeroids⁴ but unlike those of modern sharks and rays², lacks any enamel-like surface tissue. The relation of this brush to skeletal radials and proteinaceous fin rays is disputed, and it has been described as either an aberrant fin⁵ or a morphological novelty^{1,6} (but such interpretations may overlap).

The spine-brush complex may have been used in swimming, defence and courtship^{1,5-7}. It has been suggested¹ that the spine-brush has histological similarities with the corpus cavernosus of the mammalian penis, and that the brush may have been erectile¹. Furthermore, the splayed, tooth-like, apical scales of an engorged brush would have matched a similar array of scales on the head (Fig. 1a), thereby resembling gaping jaws to potential predators.

We have studied well preserved spine-brush specimens from Bearsden⁷ and show that the brush is not fibrous¹. The superficially fibrous appearance is created by a monolayer of hollow rods surrounding a substantial mineralized keel (Fig. 1a,c). The keel, the surrounding rods and the baseplate consist of globular calcified cartilage (GCC) with characteristic concentric growth lines (Fig. 1d). Exposed areas of GCC on the baseplate and elsewhere lack the tessellating polygonal units that form the distinctive surface texture of chondrichthyan prismatic cartilage. A remarkably high content of crystal fibre bundles form a meshwork in the GCC matrix of the rods and exposed parts of the baseplate (Fig. 1c).

The spine itself consists of osteonal dentine surrounded by acellular bone (Fig. 1b). A further, thin, acellular bone layer of variable thickness coats the baseplate GCC and the brush base (Fig. 1a, e). The distribution, thickness and extent of this coating varies between specimens.

The degree and topographic extent of brush mineralization shows that it was not erectile, whereas the mammalian corpus cavernosus is inflatable. The presence of such a large endoskeletal body of GCC in a jawed vertebrate is surprising. This form of calcified cartilage is otherwise known in extinct primitive vertebrates such as *Eriptychius*, osteostracans and placoderms^{3,8}. Prismatic cartilage, a fundamental chondrichthyan character⁸, probably represents an evolutionary derivative of GCC⁹. This indicates that primitive chondrichthyans may have emerged with characteristic shark-like scales¹⁰ but with 'preprismatic' GCC-based endoskeletons. The club-like extremities of the 'claspers' (male chondrichthyan intromittent organs) in the Bearsden *Stethacanthus* also consist of non-prismatic GCC.

Another possibility is that the brush acellular bone coating (Fig. 1e), although

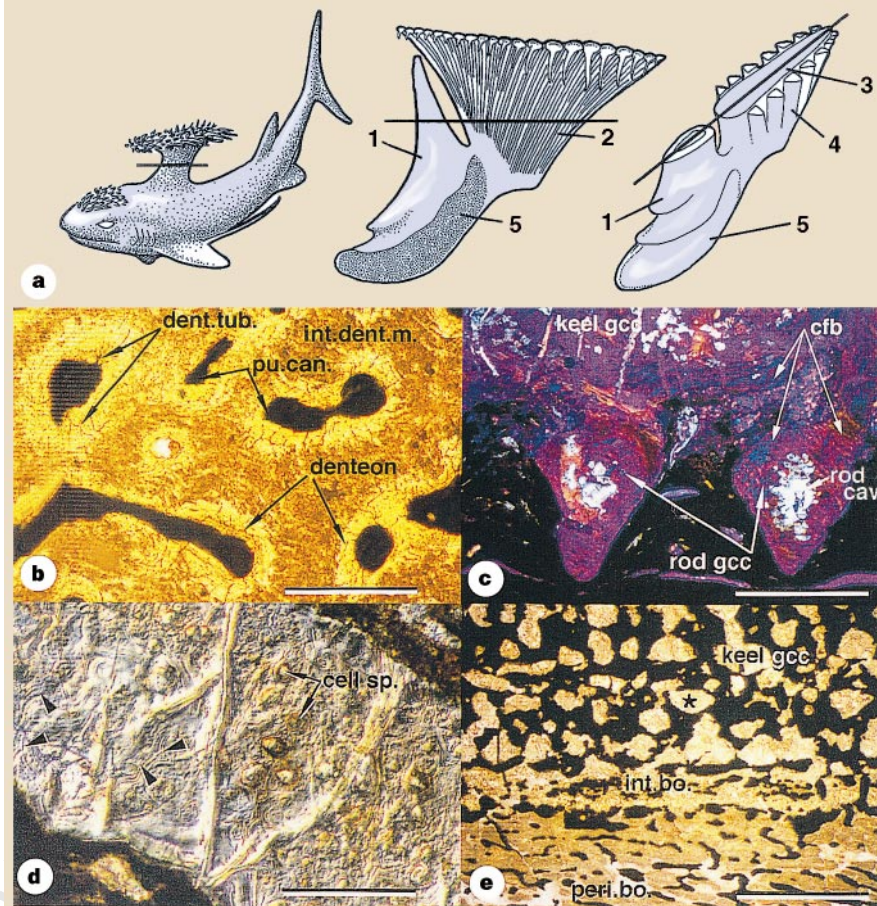


Figure 1 Structure and histology of the 'spine-brush complex' in the primitive shark *Stethacanthus*. **a**, Left to right: sketch reconstructions of *Stethacanthus*; side view of skeletal 'spine-brush' complex (teeth-like apical scales have been depressed to illustrate the length of the complex); horizontal section through spine and brush (the section level is indicated by a horizontal line in the left two diagrams). 1, Fin spine; 2, brush; 3, keel; 4, rod; 5, baseplate. The stippled area of the baseplate indicates exposed globular calcified cartilage; the non-stippled area indicates bone coating. Proportions are based on specimen HM V8246; anterior is to the left. **b**, Osteodentine (from specimen NMS 1981.63.24) from the central part of the spine shows poorly mineralized interdentinal matrix (int. dent. m.) around vascular spaces (pulp canals, pu. can.), each infilled by a dentiteon. Dentine tubules (dent. tub.) radiate from the dentiteon pulp canals. Scale bar, 160 μ m. **c**, Mineralized cartilage (from specimen NMS 1981.63.22A) of two of the rods with central hollows (rod cavity, rod cav) and the internal keel, in which blue and yellow show opposite orientations of the crystal fibre bundles (cfb), indicating strong attachment of fibres to surrounding connective and musculoskeletal tissue. The cartilage of both rod and keel is mineralized in a globular pattern (rod gcc and keel gcc). Scale bar, 360 μ m. **d**, Cartilage from the baseplate/base of the brush keel (from specimen NMS 1981.63.22A) region. The characteristic concentric growth lines based on the centres of the globules (some with central spaces, cell sp., possibly occupied by cells in life) and the waves of progressive calcification growth lines formed after fusion of the globules (Liesegang lines, arrowheads) can be seen. Scale bar, 100 μ m. **e**, A field from one side of the spine baseplate and brush base region (from specimen NMS 1981.63.22A) in transmitted light (enhanced with polarized light), showing the inner globular calcified cartilage (keel gcc) merging with an outer layer of compact acellular bone (peri. bo.). A poorly mineralized junctional tissue (int. bo.) lies between them. Asterisk, field of view shown in **d**. Scale bar, 900 μ m. HM, Hunterian Museum, University of Glasgow; NMS, National Museums of Scotland, Edinburgh.

curiously spongy, represents the earliest record of perichondral, and therefore endoskeletal, bone in a primitive chondrichthyan. With few exceptions¹¹, the absence of endoskeletal bone is seen as a defining feature of extant chondrichthyans⁸, relative to its primitive presence in other jawed and fossil jawless vertebrates. This unexpected juxtaposition of bone and GCC may represent a further, persistent, primitive aspect of early chondrichthyan skeletal tissues.

Alternatively, the near-continuity of this bone layer with the fin spine could indicate an unusual distribution of dermal skeleton derived from the trunk or cranial neural crest¹². Finally, the disputed relationship of the brush to the fins^{1,5} is resolved by regarding it as a specialized fin-baseplate extension. Its endoskeletal location, histology and absence of fin radials support this idea.

M. I. Coates*, S. E. K. Sequeira*,

I. J. Sansom†, M. M. Smith‡

*Biology Department,

University College London,

London WC1E 6BT, UK

e-mail: m.coates@ucl.ac.uk

†School of Earth Sciences,

University of Birmingham,

Birmingham B15 2TT, UK

‡Department of Craniofacial Development,

School of Dentistry,

UMDS Guy's Hospital Campus,

London SE1 9RT, UK

1. Zangerl, R. J. *Vert. Paleontol.* **4**, 372–378 (1984).
2. Maisey, J. G. *Zool. J. Linn. Soc.* **66**, 161–183 (1979).
3. Smith, M. M. & Hall, B. K. *Biol. Rev.* **65**, 277–373 (1990).
4. Patterson, C. *Phil. Trans. R. Soc. Lond. B* **249**, 101–205 (1965).
5. Lund, R. *Ann. Carnegie Mus.* **45**, 161–178 (1974).
6. Williams, M. E. *Paleontographica* **190**, 83–158 (1985).
7. Wood, S. P. *Nature* **297**, 574–577 (1982).
8. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, 1996).
9. Örvig, T. *Arkiv Zool.* **2**, 321–454 (1951).
10. Sansom, I. J. *et al. Nature* **379**, 628–630 (1996).
11. Peignoux-Deville, J., Lallier, F. & Vidal, B. *Cell Tissue Res.* **222**, 605–614 (1982).
12. Smith, M. M. *et al. Proc. R. Soc. Lond. B* **256**, 137–145 (1994).

Anoxic bioremediation of hydrocarbons

The contamination of soils and sediments by petroleum is a matter of international concern because of the toxicity and refractory character of the aromatic components in the absence of oxygen¹. Gaseous oxygen can be injected into the anaerobic zone of a contaminated environment² to stimulate biodegradation, but this is costly and inefficient. Other more soluble electron acceptors, such as nitrate or sulphate, can be used instead, but oxidation is slow and hydrocarbon degradation is incomplete³. Here we describe how chlorite dismutation by perchlorate-reducing bacteria can be used as an alternative source of oxygen for degrading contaminants. This dismutation

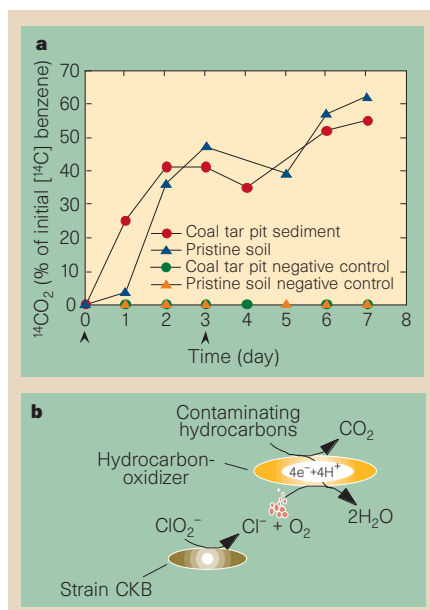


Figure 1 Stimulation of aromatic hydrocarbon oxidation by chlorite dismutation. **a**, Oxidation of [¹⁴C]-benzene to ¹⁴CO₂ in anoxic contaminated and pristine soil and sediment samples amended with chlorite and inoculated with strain CKB. Arrowheads, points of addition of 1 mM chlorite. **b**, Mechanism by which chlorite dismutation might stimulate hydrocarbon oxidation in anoxic sediments.

of chlorite into molecular oxygen and chloride is an intermediate step in the microbial reduction of perchlorate or chlorate⁴.

As part of a study on microbial perchlorate reduction, we isolated a new microorganism, strain CKB, which grows anaerobically by reducing perchlorate or chlorate, from waste sludge from a paper mill in Pennsylvania. When strain CKB is inoculated with chlorite into anoxic, petroleum-contaminated soil samples, [¹⁴C]-benzene is rapidly oxidized to ¹⁴CO₂, with about 40% of the original ¹⁴C being recovered in this form after two days of incubation (Fig. 1a). If the sediments are further amended with 1 mM chlorite on day 3, about 60% of the ¹⁴C can be recovered as ¹⁴CO₂ by day 6 (Fig. 1a). No ¹⁴CO₂ is produced in samples that are not amended with chlorite or strain CKB.

Similar results are obtained with anoxic soil samples that have had no previous exposure to hydrocarbons (Fig. 1a). However, there is a slight lag phase of 24 hours, which is consistent with adaptation to benzene by the microbial population. The stimulatory effects are not significantly altered when lower chlorite concentrations are used: 1 μM chlorite resulted in more than half the level of benzene degradation observed with 1 mM chlorite.

This stimulatory effect also occurs in defined mixed cultures in the absence of soil. When an anaerobic, washed-cell suspension of strain CKB is combined with the aerobic hydrocarbon-oxidizing *Pseudomonas*

strain JS-150 and amended with chlorite under anaerobic conditions, [¹⁴C]naphthalene is rapidly oxidized to ¹⁴CO₂ (data not shown). No ¹⁴CO₂ is produced if either of the organisms or chlorite is omitted, unless O₂ is added to the headspace. The degradation of naphthalene is therefore directly dependent on the combined presence of both strain CKB and chlorite.

Because strain CKB cannot degrade aromatic hydrocarbons in pure culture, we considered the possibility that the stimulation of hydrocarbon degradation could be the result of chlorite dismutation into chloride and O₂ by strain CKB. The resultant O₂ is used by indigenous, aerobic hydrocarbon-oxidizing bacteria that are otherwise inhibited by the anoxic condition of the soil (Fig. 1b). In support of this, when chlorite is amended to an anaerobic, washed whole-cell suspension of strain CKB, O₂ is rapidly and proportionally produced. There is no O₂ production if the cells are omitted or killed by heat.

Our results show that the dismutation of chlorite by perchlorate-reducing bacteria in anaerobic environments can produce extracellular O₂. This O₂ can be used by hydrocarbon-oxidizing bacteria to degrade hydrocarbons, such as benzene, which is a particularly important environmental contaminant owing to its toxicity and relative solubility. Little is known about perchlorate-reducing bacteria, although they are ubiquitous in a wide range of environments, including pristine soils and petroleum-contaminated sediments⁵.

High concentrations of chlorite may be toxic to many microbial species, but our results indicate that significant degradation of hydrocarbons can be stimulated at chlorite concentrations (1 μM is 90 μg per litre) well below the limits imposed by the World Health Organization (200 μg per litre) and the US Environmental Protection Agency (1 mg per litre)^{6,7}. As a bioremediative strategy, the application of chlorite dismutation to stimulate hydrocarbon oxidation in contaminated environments offers a new alternative to other injection processes.

John D. Coates, Royce A. Bruce,

John D. Haddock

Department of Microbiology,

Southern Illinois University,

Carbondale, Illinois 62901, USA

e-mail: jcoates@micro.siu.edu

1. Anderson, R. T. & Lovley, D. R. *Adv. Microbiol. Ecol.* **15**, 289–350 (1997).
2. Crocetti, C. A., Head, C. L. & Ricciardelli, A. J. *Aeration-Enhanced Bioremediation of Oil-Contaminated Soils: A Laboratory Treatability Study* (National Groundwater Association, Houston, 1992).
3. Coates, J. D., Anderson, R. T. & Lovley, D. R. *Appl. Environ. Microbiol.* **62**, 1099–1101 (1996).
4. Rikken, G., Kroon, A. & van Ginkel, C. *Appl. Microbiol. Biotechnol.* **45**, 420–426 (1996).
5. Michaelidou, U., Bruce, R., Achenbach, L. & Coates, J. D. in *Abstr. 98th Gen. Meeting Am. Soc. Microbiol.* 313 (American Society for Microbiology, Atlanta, 1998).
6. Foundation for Water Research *Research Report FR0390* (1993).
7. US Environmental Protection Agency *Fed. Reg.* **59**, 6331–6444 (1994).

Of snowflakes, symmetries and shells

The Self-Made Tapestry: Pattern Formation in Nature

by Philip Ball

Oxford University Press: 1998. 287 pp.

£18.99, \$37.50

Edward Cox

One of the great and unexpected pleasures of the play *Breaking the Code* by Hugh White more occurred when Derek Jacobi, playing Alan Turing, faced the audience and talked about spatial patterns in fir cones, and how they can be described by the Fibonacci series. When the play first opened at the Kennedy Center, my wife and I immediately bought tickets and took the train to Washington, figuring it would close after one or two performances. We were wrong. The audience was spellbound by Jacobi's soliloquies on mathematical logic, computing and pattern formation in animals and plants. The fact that *Breaking the Code* then went on to Broadway and enjoyed great success is still a bit of a puzzle — are there really enough biologists and mathematicians around to sustain a good Broadway run?

Part of the key to understanding the play's success, I think, lies in everyone's interest in the origins of symmetries and patterns in nature. It may even be that our visual system, with its remarkable abilities as a pattern detector, is at the root of this interest. It goes quite deep in another sense too, a need to discover first causes. We look for patterns in space and time, of course, but we also try to understand where they come from — how radial symmetries emerge from a spherical egg, why honeycombs are hexagonal, how radiolarians create beautifully patterned exoskeletons, how snowflakes form, the universal patterns apparent in mountainscapes and why soap bubbles pack the way they do.

In his new book, Philip Ball talks about these questions at engaging length and with a balance, thoroughness and simplicity that I found remarkable. In ten chapters he moves from bubble and foam structure, subjects that can be understood by appeal to simple geometry and physics, to travelling and standing waves used by living systems to set up coordinates, and then on to fluid dynamics, community ecology, and finally a chapter on physical principles, the best introductory chapter on this subject I have read.

In a sense, the first part of the book is divided into topics inspired by D'Arcy Thompson's *On Growth and Form* and by Turing's remarkable paper, "The chemical basis of morphogenesis". Although both authors searched for physical explanations, Thompson's took the form of mathematical description and mechanical principle, while for Turing the language of mathematics was



In formation: "self-made" patterns induced in a bacterial colony by chemotaxis.

used to describe the emergence of patterns from apparently featureless initial conditions. Later in the book, Ball turns to contemporary discussions of branching patterns (snowflakes, crack propagation in solids, viscous fingering, diffusion-limited aggregation), fluid flow, sand-piles and self-organized criticality.

There are very few patterns in biology, physics and chemistry that Ball fails to examine carefully, illustrating along the way principles of self-organization and thermodynamics that work at all levels and length scales. There are particularly fine expositions of bubble and foam structure, which he introduces by discussing D'Arcy Thompson's views on honeycomb symmetries.

Ball's style is historical and comparative, illustrating the twists and turns that lead from observation to understanding. This makes for fascinating reading — just when I thought I understood why bees are such excellent geometers, I was given a new insight which, it turned out, also had its problems. Perhaps the true explanation has now arrived with the work of D. Weaire and R. Phelan on bubble packing (*Nature* 367, 123; 1994).

Ball describes the essential features of Turing's contribution to the genesis of spatial patterns, which has proved to be a very rich source of ideas in chemistry and biology, illustrating by example how local autocatalysis and long-range inhibition can yield such amazing variety. His examples are all aptly chosen: from biology we have sea-shell and animal coat patterns, the eyespots on butterfly wings, spiral waves in *Dictyostelium* territories and chemotaxing bacteria on agar surfaces; and from physics, wave propagation on catalytic surfaces, Liesegang rings in tubes and flow patterns in cylindrical flames. He gives the reader a good feeling for

why order can emerge from these disparate systems, which vary enormously in scale and physical mechanism.

At times the going is tough. The Belousov-Zhabotinsky reaction can finally only be understood by quantitative arguments, I think. No amount of appeal to metaphor seems to work really well. Still, as one who has failed many times at teaching this reaction, I sympathize, and see this as a generic problem when discussing complex systems. When all is said and done, in some systems and circumstances, the interested reader must model complexity by coming to grips with the mathematics of the system. Still, it is remarkable, I think, that Ball manages more often than not to do so well without formal arguments.

Many authors who cover grand and complicated topics do so by sacrificing accuracy in their search for simple language. Ball's approach is to write carefully, be sceptical and non-polemical about data and their interpretation, and give credit to heterodox views where credit is due. In sum, he is critical of the literature and its interpretation, perhaps, as he says in the introduction, because his "years at *Nature* magazine have exposed me to too many amazing discoveries that vanish like morning mist under close scrutiny".

Since the appeal of patterns in nature has such a strong aesthetic component, it is a pleasure to report that this book is handsomely and accurately printed and produced. It contains many wonderful illustrations, both in colour and in black and white, and very few errors, factual or typographical. □

Edward Cox is in the Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA.

Life, the Universe and the single mollusc

Civilization and the Limpet

by Martin Wells

Perseus Books: 1998. 224 pp. \$22, £15.50

John Shepherd

Do you know the link between the sexual habits of the slipper limpet and a Roman aqueduct? Or why slipper limpets are sufficiently remarkable to be favoured to the specific name *Crepidula fornicata*? Have you ever wondered how ordinary limpets find their way home after a busy day cleaning algae off the rocks? Do you know the best way to catch an octopus? Or why whales don't get the bends? The answers to these and many other intriguing questions can be found in this



Safely grazing: limpets' navigational powers will help them to find their way home.

engagingly haphazard collection of essays.

The author, Martin Wells, is a marine biologist, an expert on cephalopods and a yachtsman. He originally conceived the book as a journal of a voyage, but his enthusiasm for his subject has led him astray, and we can be grateful for that. Life is, after all, what happens to you while you are making other plans. Each essay is based on some aspect of the life of an animal in the sea, but develops fugue-like to encompass observations on life, the Universe and even, as the title promises, civilization. The tone is set, and Wells's philosophy of life begins to beguile, on the very first page of his preface, when he remarks that "some poor people work for all their lives, and all they ever make is money".

He develops his discussion on the navigational powers of limpets — through a discussion on the perceptual powers of octopuses in two and three dimensions — to explain how these are affected by the limpets' lack of bones and joints. And thence to the ability of vertebrates to build cars and computers, and to bash the ozone layer. In a similar vein, one chapter begins with an uncle who was interested in lugworms and defaecation, and develops into a discussion of biological clocks and free will. A comment on the sheltered life led by physicists, and its limiting effect on their powers of imagination, leads via the effects of entropy (and how to contain them using a can of WD-40) to your chances of surviving the next mass extinction.

Wells's own speciality gets its share of the action, too, with accounts of the swimming and depth-control abilities of squid, and the locomotive and reproductive adaptations of *Nautilus*, its ability to tolerate low oxygen levels, and how this may be relevant to the success of ammonites in the remote past. In a final chapter, "Does science have to be useful?", he mounts a spirited attack on the utilitarian orthodoxy of our time, asserting the claim of science (like the arts) to be supported not least because it amuses, entertains and enriches our lives. Along the way he claims that biologists are more realistic than 'hard' scientists in their judgements because they are used to dealing with the messiness of the natural world. I think physicists working as meteorologists, among others, may disagree with that.

My only complaint is that there are no illustrations to add further spice to this rich

mixture. If you need an antidote to the world of targets, objectives, milestones and deliverables, I heartily recommend this book. A scientist who is willing to say that his studies of *Nautilus* give him an excuse "to swan off to the Pacific, and there spend a quite unnecessary proportion of my time underwater" should be cherished, for this is not politically correct, and such a scientist may become an endangered species. Read this and recapture the pleasure of mucking about in rock pools, just for fun. But spare a thought for the limpet, because "a limpet has nothing, or next to nothing, to fantasize about". □

John Shepherd is at the Southampton Oceanography Centre, University of Southampton, Empress Dock, Southampton SO14 3ZH, UK.

Africa's mistletoes ripe for the picking

Mistletoes of Africa

by Roger Polhill and Delbert Wiens
Royal Botanic Gardens, Kew: 1998.
370 pp. £70

Jeremy Midgley

Mistletoes are rootless plants that plug into the 'plumbing' of their host plant's stem to steal water, dissolved minerals and, to varying degrees, organic products. They are a unique yet global group, full of fascinating attributes. But until now there has been no comprehensive book on the African species. Roger Polhill and Delbert Wiens' *Mistletoes of Africa* gives an up-to-date summary of mistletoe biology, keys to the taxa and some biogeography, as well as describing their haustoria (by which they attach to the host) and flower structure in relation to opening, and providing appetizers for further work, for which they note a strong need.

Mistletoes may be in danger of extinction for several reasons. They are probably prone to excessive herbivory by large African animals like elephants. It has even been suggested that the absence of large animals, such as giraffe, in some reserves leads to a build up of mistletoes on their preferred hosts, thorny acacias. This increase in mistletoes attracts fruit-eating birds which in turn produces an influx of seeds of fleshy-fruited, broad-leaf species that become established beneath the acacia. Eventually, acacia savanna is replaced by dense, broad-leaf woodlands, the phenomenon of bush encroachment. This is perhaps an exaggeration but certainly more work is needed on the relationships between large African mammals and mistletoes.

Some mistletoes may be vulnerable because of their obligate relationships with mutualists. Dioecious fleshy-fruited species, like *Viscum continuum*, depend on animals

for both pollination and dispersal. People are also potentially disastrous for mistletoes. 'Wood roses', shaped from haustoria of savanna mistletoes, are common in the curio markets of South Africa. They are carved from mistletoes that can be more than 50 years old; excessive use may not be sustainable.

A largely African aspect of mistletoe ecology is their occurrence on succulents, where some ramify within the tissues of their hosts. They reach their most advanced and reduced stage in *V. minimum* of the eastern Cape, a really weird plant. Its shoot (less than 3 millimetres long, see below) is far smaller than the berry it produces; in fact, the whole plant looks like its host plant's fruit.

Non-endophytic mistletoes are almost globally restricted to arid regions of Africa and their physiology is intriguing. We do not know whether they use the stored water or the xylem stream of their host plant. Mistletoes open their stomata and fix carbon during the day; succulents, on the other hand, open their stomata at night — an uneasy relationship. Further physiological pieces of this fascinating jigsaw are needed.

Host relations also need research. Polhill and Wiens suggest, for example, that host specialization is more frequent when potential host diversity is low. Some species are even restricted to parasitizing other mistletoes (for example, the epiparasitism of *V. goetzii* and *V. loranthicola*). Others are single dots on the African map, such as *V. grandicaule*, the largest *Viscum*, which only grows on fig trees on the island of Pagalu, off Gabon.

African mistletoes are similar to those found elsewhere in the world as well as having many unique characteristics. With the publication of this book, work on African mistletoes can now safely enter an "after-description phase". □

Jeremy Midgley is in the Botany Department, University of Cape Town, P Bag Rondebosch, 7701 South Africa.



Decoratively minimal: *Viscum minimum* is fruit and little else.



Quantitative test of a microscopic mechanism of high-temperature superconductivity

Eugene Demler*† & Shou-Cheng Zhang*

* Department of Physics, Stanford University, Stanford, California 94305, USA

One of the main challenges to theoretical attempts to understand the microscopic mechanism of high-transition-temperature (high- T_c) superconductivity is to account quantitatively for the superconducting condensation energy, the energy by which the normal state differs from the superconducting state^{1–6}. A microscopic model commonly used to describe the superconducting copper oxides, the t - J model⁷, is thought to capture the essential physics of the phenomenon: the interplay between the electrons' kinetic energy and their antiferromagnetic exchange interaction. Within the t - J model the condensation energy can be related to the change in the dynamical spin structure between the superconducting and the normal states⁸. Here we propose a microscopic mechanism for the condensation energy of high- T_c superconductors. Within this mechanism, the appearance of a resonance in the superconducting state^{9–13} enables the antiferromagnetic exchange energy in this state to be lowered relative to the normal state. We show that the intensity of the resonant neutron-scattering peak observed previously in $\text{YBa}_2\text{Cu}_3\text{O}_7$ when it undergoes the transition to the superconducting state^{14–16} is in quantitative agreement with the condensation energy of these materials^{2,3}.

One reason for the absence of a generally accepted microscopic explanation of high- T_c superconductivity (HTSC) is the lack of quantitatively precise definition of a 'mechanism'. In the traditional BCS¹ superconductors, the phonon mechanism is quantitatively verified by isotope substitution and single particle tunnelling experiments. It is a widely held view that HTSC involves strong electronic interactions rather than phonon-mediated electron pairing. To date, many different 'mechanisms' have been proposed, but very few quantitatively falsifiable predictions have been checked against experimental data.

A system undergoes a transition from the normal to the superconducting state because it can lower the total free energy. The condensation energy E_C , defined as the energy difference between the normal state, extrapolated to zero temperature, (E_N) and the superconducting state (E_S), can be directly measured through the thermodynamical critical field H_c by the following relation¹:

$$E_C = E_N - E_S = \frac{H_c^2}{8\pi} V_0 \quad (1)$$

Here we choose $V_0 = a \times b \times c$ to be the volume of a unit cell (in the high- T_c copper oxides, $a = b$ with reasonably good accuracy) and define the energies inside one unit cell. Alternatively, the condensation energy can also be found by integrating the difference in specific heats in the (extrapolated) normal and superconducting states from $T = 0$ to the superconducting transition temperature T_c (refs 2, 3).

On the other hand, the condensation energy can, in principle, be computed from a microscopic hamiltonian. Ideally, one would like to start from the full hamiltonian involving electronic kinetic energy, the Coulomb interaction energy and the electron lattice coupling energy. Chester⁴ has studied the condensation energy of such a hamiltonian and showed how various contributions to the condensation energy

could in principle be determined experimentally. Among many theories of HTSC, only the interlayer tunnelling mechanism⁵ made a quantitatively falsifiable prediction about the condensation energy. More recently, Leggett⁶ showed that the change in the interlayer Coulomb energy can be directly related to the frequency and momentum integral of the dynamic charge susceptibility, a quantity which can be measured experimentally. He argued that a microscopic theory of HTSC should quantitatively predict the frequency and the momentum range at which the change of the dynamical susceptibility occurs, and quantitatively account for the condensation energy.

But the energy scale involved in HTSC is much smaller than that of the Coulomb interaction. To understand the mechanism for HTSC, it is useful to start with an effective hamiltonian which correctly describes the basic physics below the Coulomb energy scale. Recently, Scalapino and White⁸ initiated the investigation of condensation energy in the t - J model, which is defined by:

$$H = -t \sum_{\langle ij \rangle} c_{i\sigma}^\dagger c_{j\sigma} + J \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j \quad (2)$$

Here $c_{i\sigma}^\dagger$ is the electron creation operator on site i with spin σ , $\mathbf{S}_i$ is the electron spin operator, $\langle ij \rangle$ denotes a pair of near-neighbour sites and the hamiltonian does not act on the space of doubly occupied sites. Equation (2) describes the hamiltonian of one layer only. In a bilayer system like $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, there are two layers per unit cell, with additional antiferromagnetic coupling $J_\perp$ between the layers. If a system described by this hamiltonian undergoes a superconducting transition away from half-filling of the electron band, the kinetic energy E_t is expected to increase slightly, but the decrease in the exchange energy E_J may be significantly larger, so that a superconducting ground state is realized.

In a tight-binding model, the mean kinetic energy E_t can be expressed as a frequency integral of the optical conductivity $\sigma(\omega)$ (refs 17–19). Then the contribution of E_t to the condensation energy can be measured by comparing $\sigma(\omega)$ in the normal and the superconducting states. In the low-frequency regime, the Drude peak in the normal state collapses into a δ -function peak in the superconducting state, usually with conserved weight²⁰. Therefore, the change in the kinetic energy has to be determined from $\sigma(\omega)$ in the higher-frequency range.

Scalapino and White⁸ made the insightful observation that the change in the J term in equation (2) can also be directly expressed as a frequency and momentum integral of the dynamic spin structure factor $S(q, \omega)$. For bilayer materials such as $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, it is convenient to separate the even and odd parts of $S(q, \omega)$ with respect to bilayer interchange within a unit cell:

$$S(q, \omega) = S^{\text{even}}(q, \omega) \cos^2\left(\frac{q_z d_{\text{Cu}}}{2}\right) + S^{\text{odd}}(q, \omega) \sin^2\left(\frac{q_z d_{\text{Cu}}}{2}\right) \quad (3)$$

Here d_{Cu} is the distance between nearest-neighbour Cu–O planes, so S^{even} and S^{odd} describe the in-phase and out-of-phase spin fluctuations in the two planes constituting one bi-layer. Then $\Delta E_J = E_J^N - E_J^S$, defined as a difference between exchange energies in the normal (E_J^N) and the superconducting (E_J^S) states, may be obtained (per unit cell) as:

$$\begin{aligned} \Delta E_J = & \frac{3}{2} \left(\frac{a}{2\pi} \right)^2 \int_{-\pi/a}^{\pi/a} d^2 q \int_0^\infty \frac{d(\hbar\omega)}{\pi} \{ [S_N^{\text{even}}(q, \omega) - S_S^{\text{even}}(q, \omega)] \\ & \times \left[\frac{1}{2} J(\cos(q_x a) + \cos(q_y a)) + \frac{1}{4} J_\perp \right] \\ & + [S_N^{\text{odd}}(q, \omega) - S_S^{\text{odd}}(q, \omega)] \\ & \times \left[\frac{1}{2} J(\cos(q_x a) + \cos(q_y a)) - \frac{1}{4} J_\perp \right] \} \end{aligned} \quad (4)$$

Here $q = (q_x, q_y)$ is a two-dimensional in-plane momentum and

† Present address: Institute for Theoretical Physics, University of California, Santa Barbara, California 93106-4030, USA.

$S_N^{\text{even/odd}}$ and $S_S^{\text{even/odd}}$ correspond to the dynamic spin structure factors in the normal and the superconducting states, respectively. It is known^{21,22} that $J_\perp$ is much smaller than J , so the $J_\perp$ term in equation (4) may be safely neglected. Scalapino and White pointed out⁸ that as $S(q, \omega)$ can be measured directly in neutron-scattering experiments, ΔE_J can therefore in principle be measured. If ΔE_J turns out to be much smaller than the experimentally measured condensation energy E_c , this gives a convincing way to rule out any mechanism of high- T_c superconductivity based on the t - J model. On the other hand, if ΔE_J turns out to be greater than E_c one would have identified an important driving force for superconductivity, and can quantitatively test any theoretical mechanism of HTSC based on the t - J model, which should predict where in frequency and momentum space the change of $S(q, \omega)$ occurs.

However, one should apply this line of reasoning with extreme care. The quantity $S_N(q, \omega)$ in equation (4) is not the normal state spin structure factor above T_c , but rather an extrapolated normal state quantity at $T = 0$. Experimentally, one has to carefully identify features in $S(q, \omega)$ which change abruptly at T_c . Theoretically, one has to identify contributions to $S(q, \omega)$ which are absent in the normal state and only present in the superconducting state, so that the difference in equation (4) can be rigorously defined. A resonant neutron peak has been discovered in the family of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ superconductors^{14–16}. In the optimally doped $\text{YBa}_2\text{Cu}_3\text{O}_7$ superconductor with $T_c = 91$ K, the resonance peak occurs in the spin scattering channel at energy $\omega_0 = 41$ meV. It is centred around momentum $Q = (\pi/a, \pi/a)$ and it occurs in the odd channel with respect to bilayer interchange. Perhaps the most remarkable feature of the resonance peak is its onset at the superconducting transition temperature T_c . Above T_c , there is no experimentally identifiable spectral weight in this frequency and momentum range. This experiment is therefore ideally suited for analysing the contribution to the condensation energy due to the exchange interaction.

We constructed a theory⁹ of the neutron resonance peak by first identifying a resonance in the spin-triplet particle-particle (p-p) channel defined by the following operator:

$$\pi^\dagger = \sum_k (\cos k_x - \cos k_y) c_{k+Q}^\dagger c_{-k}^\dagger \quad (5)$$

where $c_{k_i}^\dagger$ creates a spin-up electron at momentum $k = (k_x, k_y)$.

In a series of exact diagonalization studies on the Hubbard and t - J models, Meixner, Hanke and the two of us¹⁰, and Eder, Hanke and one of us (S.-C.Z.)¹¹, demonstrated the existence of such a triplet p-p resonance for all doping ranges. A triplet p-p resonance cannot be detected in neutron-scattering experiments on the normal state. However, below T_c , the d -wave superconducting order parameter mixes a p-p resonance into the dynamical spin structure factor, leading to a sharply defined collective mode in the triplet particle-hole (p-h) channel (the so-called π -resonance). The contribution from the p-p resonance to the p-h channel below T_c is given^{9,12} by:

$$\int d(\hbar\omega) S(q, \omega) = \frac{2\langle\Delta_d\rangle^2}{1-n} \quad (6)$$

where $\langle\Delta_d\rangle$ is the dimensionless d -wave superconducting order parameter (not the energy gap) and n is the density of electrons. This simple estimate of the spectral weight of the π resonance at $q = Q$ agrees reasonably well with the absolute intensity measured by Fong *et al.*^{12,16}. On the other hand, the width of the π resonance in momentum space cannot be calculated reliably. Within the $\text{SO}(5)$ theory¹³, the π resonance has been interpreted as a pseudo-Goldstone boson of the spontaneous $\text{SO}(5)$ symmetry breaking in the superconducting state. Its existence is therefore directly correlated with the existence of the superconducting long-range order.

Because the coupling to a p-p resonance is only possible in the superconducting state, it indeed contributes to the difference in $S(q, \omega)$ as required by equation (4). We point out here that the logic leading to our explanation of the π resonance below T_c can be

reversed to provide a microscopic mechanism of HTSC. On one hand, the existence of the π resonance is a unique property of the superconducting state; on the other hand, its existence around Q lowers E_J , as we see from equation (4). Therefore, the system undergoes a superconducting transition, so that a π resonance mode can emerge, which in turn lowers E_J . In other words, no matter how antiferromagnetically correlated is the normal state, a superconducting state can always further lower the exchange energy by coupling a p-p resonance into the dynamic spin structure factor $S(q, \omega)$. Our mechanism for HTSC makes a quantitatively precise prediction which can be checked experimentally. Once we assume that the HTSC is dominantly driven by the emergence of a new collective mode, the π resonance, this mechanism predicts that the condensation energy is determined by the zero-temperature spectral weight of the π resonance, and the energy saving occurs precisely in the frequency and momentum range of the π resonance.

As our mechanism predicts an additional spectral weight around $q = Q$ in the superconducting state, one may wonder what happens to the spectral sum rule, which states that $S(q, \omega)$ integrated over both q and ω (without the $f(q) = \cos(q_x a) + \cos(q_y b)$ factor in equation (4)) should be a constant, independent of the nature of the ground state. Our additional spectral weight should, of course, be compensated by a depletion of spectral weights from other parts in the momentum space. However, because $f(q)$ has an absolute minimum at $q = Q$, their contributions to (4) cannot completely cancel the π resonance contribution. In fact, depletion of the spectral weight near $q = 0$ would lead to additional enhancement of the condensation energy.

Both the condensation energy and the spectral weight of the neutron resonance peak have been measured experimentally, so it is straightforward to check this prediction. Within Landau–Ginzburg theory, $H_c = \Phi_0(2\sqrt{2\pi\xi\lambda})$, where $\Phi_0 = hc/2e$ is the superconducting flux quantum, ξ is the coherence length and λ is the London penetration depth. In $\text{YBa}_2\text{Cu}_3\text{O}_7$, these two length scales are determined to be in the range of $\xi = 12$ – 20 Å and $\lambda = 1,300$ – $1,500$ Å. Using $a = b = 3.85$ Å and $c = 11.63$ Å, this gives a condensation energy of $E_c = 3.3$ – 12 K per unit cell. On the other hand, E_c can also be measured directly in specific-heat experiments. Loram *et al.* reported (see page 251 of ref. 2 and page 247 of ref. 3) a value of $E_c = 6$ K per unit cell for $\text{YBa}_2\text{Cu}_3\text{O}_7$. We now turn to the experimental measurement of ΔE_J given by equation (4). In practice, it is very hard to measure the change of $S(q, \omega)$ throughout the frequency and momentum range. However, because our mechanism predicts that most of the condensation energy is saved in the range of $\omega_0 = 41$ meV and a two-dimensional momentum transfer $Q = (\pi/a, \pi/a)$, we can analyse the neutron scattering data by assuming that the contributions of the other parts do not change significantly at T_c . Indeed, for frequencies above the resonance the neutron-scattering intensity does not change as the system goes from normal to superconducting. For frequencies below and in the resonance range, the normal state intensity $S_N(q, \omega)$ is below the experimental sensitivity limit in both even and odd channels. Because of the sum rule discussed above we may not assume that it vanishes identically, so we make an assumption that in this frequency range the spectral weight $S_N(q, \omega)$ is spread uniformly in momentum space and therefore does not contribute to equation (4). Below T_c , a neutron resonance peak emerges in the odd channel, and $S_S^{\text{odd}}(q, \omega)$ has been measured in absolute units in the range of the resonance. The dimensionless quantity $\int d(\hbar\omega) S_S^{\text{odd}}(Q, \omega)$ is measured to be 0.52 at $T = 10$ K by Fong *et al.*¹⁶. (A theoretical estimate based on references 9 and 12 gives 0.32 for this value.) The resonance has a lorentzian profile centred at $q = Q$ with a width $\kappa_{2D} = 0.23$ Å^{−1} (ref. 16), so the two-dimensional momentum integral can be easily estimated.

$$\Delta E_J = \frac{3}{2} \pi \left(\frac{a}{2} \kappa_{2D} \right)^2 \frac{1}{2} \frac{0.52}{\pi} 2J = 0.016J \quad (7)$$

Taking $J = 100$ meV we obtain $\Delta E_J = 18$ K. The actual condensa-

tion energy should be smaller than ΔE_f as the kinetic energy is expected to increase below the superconducting state. Taking this into account, we see that the spectral weight of the π resonance peak can quantitatively account for the condensation energy measured from H_c and specific-heat experiments, in reasonable agreement with the prediction based on our mechanism.

We now compare our mechanism for HTSC with the spin fluctuation pairing mechanism^{23–25}. All these mechanisms are based on the antiferromagnetic exchange interaction E_f . Our analysis of the neutron data confirms that a large part of the condensation energy indeed arises from such an exchange interaction⁸. However, our mechanism differs from the spin fluctuation pairing mechanism, both conceptually and quantitatively. The idea of spin fluctuation pairing is directly borrowed from the phonon-mediated pairing in the traditional BCS superconductors. It requires sizeable antiferromagnetic spin fluctuations in the normal state, in order to pair the Fermi liquid like quasiparticles. In contrast, in our mechanism, the antiferromagnetic spin fluctuations in the normal state do not play an important role in driving superconductivity. In the superconducting state, the π resonance can be viewed as a kind of spin fluctuation in a particular frequency and momentum range. From the point of view of the $SO(5)$ theory¹³, the superconducting state is obtained from the antiferromagnetic state by a rotation, so it is natural to expect it to have more antiferromagnetic correlation than the normal state. The crucial conceptual difference between these two mechanisms therefore lies in the fact that normal state antiferromagnetic spin fluctuations are not required in our mechanism. This conceptual difference has a direct quantitative consequence. From equation (4), we see that the presence of spin fluctuation spectral weight in the normal state contributes negatively to the condensation energy. It is not clear why the spin fluctuation model would predict a change in the frequency-integrated weight of $S(q, \omega)$ near $q = Q$. In fact, most theories of the neutron resonance peak based on the spin fluctuation models^{26,27} assume a pre-existing overdamped spin fluctuation mode in the normal state and only predict a reduction of damping below T_c with essentially conserved weight.

Our mechanism provides a natural explanation for the doping dependence of the condensation energy. We attribute the condensation energy to the difference in exchange energies in superconducting and normal states. More-underdoped materials have considerably more antiferromagnetic correlations in the normal state, as known from neutron scattering²⁸. Although in the superconducting state the weight of the resonance is enhanced²⁹, the difference between the normal and the superconducting exchange energies becomes smaller. This is consistent with the results of Loram *et al.*³⁰ who find a decrease in the condensation energy with decreased doping in $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{6+x}$. Until this point we have been discussing the relation between the condensation energy and the resonant peaks in neutron scattering at zero temperature. It would be interesting to see whether this idea also works at finite temperature. Without going into details, we would like to point out a striking similarity between the temperature dependence of H_c (ref. 2) and that of the resonance intensity^{29,31} for the underdoped copper oxides: both quantities scale similarly below T_c and both have 'tails' extending above T_c , presumably arising from significant pairing fluctuations in the pseudogap regime. These phenomena await future investigations. □

Received 4 June; accepted 16 September 1998.

- Shrieffer, J. R. *Theory of Superconductivity* (Addison-Wesley, Reading, MA, 1964).
- Loram, J. *et al.* The electronic specific heat of $YBa_2(Cu_{1-x}Zn_x)_3O_7$ from 1.6 to 300 K. *Physica C* **171**, 243–256 (1990).
- Loram, J. *et al.* Electronic specific heat of $YBa_2Cu_3O_{6+x}$ *et al.* from 1.8 to 300 K. *J. Supercond.* **7**, 243–249 (1994).
- Chester, G. V. Difference between normal and superconducting states of a metal. *Phys. Rev.* **103**, 1693–1699 (1956).
- Chakravarty, S. *et al.* Interlayer tunneling and gap anisotropy in high temperature superconductors. *Science* **261**, 337–340 (1993).
- Leggett, A. Where is the energy saved in cuprate superconductivity? Preprint, Univ. Illinois at Urbana-Champaign (1998).
- Zhang, F. C. & Rice, T. M. Effective Hamiltonian for the superconducting copper-oxides. *Phys. Rev. B* **37**, 3759–3763 (1988).

- Scalapino, D. J. & White, S. The superconducting condensation energy and an antiferromagnetic exchange based pairing mechanism. Preprint cond-mat/9805075 at (<http://xxx.lanl.com>) (1998).
- Demler, E. & Zhang, S. C. Theory of resonance neutron scattering of high T_c superconductors. *Phys. Rev. Lett.* **76**, 4126–4129 (1995).
- Meixner, S. *et al.* Finite size studies on the $SO(5)$ symmetry of Hubbard model. *Phys. Rev. Lett.* **79**, 4902–4905 (1997).
- Eder, R., Hanke, W. & Zhang, S. C. Numerical evidence for $SO(5)$ symmetry and superspin multiplets in the two dimensional t - J model. *Phys. Rev. B* **57**, 13781–13789 (1998).
- Demler, E., Kohno, H. & Zhang, S. C. π excitation of the t - J model. *Phys. Rev. B* **58**, 5719–5730 (1998).
- Zhang, S. C. A unified theory based on $SO(5)$ symmetry of superconductivity and antiferromagnetism. *Science* **275**, 1089–1096 (1997).
- Mook, H. *et al.* Polarized neutron determination of the magnetic excitation in $YBa_2Cu_3O_7$. *Phys. Rev. Lett.* **70**, 3490–3493 (1993).
- Fong, H. F. *et al.* Phonon and magnetic neutron scattering at 41 meV in $YBa_2Cu_3O_7$. *Phys. Rev. Lett.* **75**, 316–319 (1995).
- Fong, H. F. *et al.* Polarized and unpolarized neutron scattering study of the dynamic spin susceptibility in $YBa_2Cu_3O_7$. *Phys. Rev. B* **54**, 6708–6720 (1996).
- Shastri, S. & Sutherland, B. Twisted boundary conditions and effective mass in Heisenberg-Ising and Hubbard rings. *Phys. Rev. Lett.* **65**, 243–246 (1990).
- Scalapino, D., White, S. & Zhang, S. C. Superfluid density and the Drude weight of the Hubbard model. *Phys. Rev. Lett.* **68**, 2830–2833 (1992).
- Chakravarty, S. Do electrons change their c -axis kinetic energy upon entering the superconducting state? Preprint cond-mat/9801025 at (<http://xxx.lanl.gov>) (1998).
- Timusk, T. & Statt, B. The pseudogap in high temperature superconductors: an experimental survey. *Rep. Prog. Phys.* (in the press).
- Reznik, D. *et al.* Direct observation of optical magnons in $YBa_2Cu_3O_{6.2}$. *Phys. Rev. B* **53**, R14741–R14744 (1996).
- Hayden, S. *et al.* High frequency spin waves in $YBa_2Cu_3O_{6.15}$. *Phys. Rev. B* **54**, R6905–R6908 (1996).
- Scalapino, D. The case for $d_{x^2-y^2}$ pairing in the cuprate superconductors. *Phys. Rep.* **250**, 329–365 (1995).
- Pines, D. Understanding high-temperature superconductivity, a progress report. *Physica B* **199–200**, 300–309 (1994).
- Shrieffer, J. R. Ward's identity and the suppression of spin fluctuation superconductivity. *J. Low Temp. Phys.* **99**, 397–402 (1995).
- Bulut, N. & Scalapino, D. Neutron scattering from a collective spin fluctuation mode in CuO_2 bilayer. *Phys. Rev. B* **53**, 5149–5152 (1996).
- Morr, D. K. & Pines, D. The resonance peak in cuprate superconductors. Preprint cond-mat/9805107 at (<http://xxx.lanl.gov>) (1998).
- Hayden, S. *et al.* Absolute measurements of the high-frequency magnetic dynamics in high- T_c superconductors. *Physica B* **241–243**, 765–772 (1998).
- Fong, H. F. *et al.* Superconductivity-induced anomalies in the spin excitation spectra of underdoped $YBa_2Cu_3O_{6+x}$. *Phys. Rev. Lett.* **78**, 713–716 (1997).
- Loram, J. W. *et al.* Superconducting and normal state energy gaps in $Y_{0.8}Ca_{0.2}Ba_2Cu_3O_{7-\delta}$ from the electronic specific heat. *Physica C* **282–287**, 1405–1406 (1997).
- Dai, P. *et al.* Magnetic dynamics in underdoped $YBa_2Cu_3O_{7-x}$: direct observation of a superconducting gap. *Phys. Rev. Lett.* **77**, 5425–5428 (1996).

Acknowledgements. We thank P. Dai, H. Fong, S. Heyden, B. Keimer, H. Mook and D. Scalapino for discussions.

Correspondence and requests for materials should be addressed to E.D. (e-mail: demler@itp.ucsb.edu).

Moisture-induced ageing in granular media and the kinetics of capillary condensation

L. Bocquet*, E. Charlaix*†, S. Ciliberto* & J. Crassous*

* Laboratoire de Physique de l'ENS de Lyon (URA CNRS 1325), 46 Allée d'Italie, 69364 Lyon Cedex, France

† DPM (UMR CNRS 5586), Université Claude Bernard-Lyon I, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

In 1773 Coulomb¹ recognized that the static properties of granular systems can be discussed in terms of the frictional properties between different layers², leading to his relationship between the angle of repose of a granular pile (θ_0) and the coefficient of static friction μ_s : $\tan \theta_0 = \mu_s$. Two centuries later, solid friction and granular media still present many puzzles. One such is that the coefficient of static friction depends on the time during which the solids remain in contact before the measurement. Here we show that this ageing effect is manifested too in the angle of repose of granular media and originates from capillary condensation of water vapour between the packed particles, leading to the formation of water bridges. By assuming that the kinetics of this process are governed by the thermally activated nucleation of bridges, we can reproduce both the time- and humidity-dependence of the ageing behaviour. Our results also clarify the kinetics of adsorption in porous media more generally.

Ageing is said to occur in systems for which the relaxation time becomes so long that it may not reach equilibrium on a laboratory timescale (typically hours). The measured properties of such a system therefore depend on the time at which the measurement is made. This is the case for solid friction: the friction coefficient depends on the time elapsed after the surfaces have come into contact^{3–5}. This dependence of μ_s on waiting time is observed with almost all materials (from paper to rocks) and is found universally to be logarithmic. Although solid friction is related to the properties of granular media, as emphasized by Coulomb's analogy, ageing in granular media has previously received little attention.

We have studied the effect of waiting time on the angle of first avalanche θ_w of a granular system of small (typically 200- μm) spherical glass beads contained in a rotating drum (Fig. 1a). We observe logarithmic ageing of the maximum static angle θ_w (Fig. 2a). This logarithmic behaviour spans more than three orders of magnitude in time, ranging from 5 seconds to more than 2 hours. Ageing was

not observed for beads with a diameter larger than 0.5 mm, except at very large humidities. It is known that humidity can exert an important influence on granular media². Addition of small quantities of wetting liquid has been shown to change enormously the repose angle of a pile^{6,7}; a brief discussion of moisture effects can even be found in Coulomb's treatise¹. We repeated our experiments at various humidities P_v/P_{sat} (P_v and P_{sat} being respectively the vapour pressure and saturated water pressure), and found humidity to be the crucial parameter controlling the ageing of θ_w : no ageing is observed at low humidity, and the magnitude of the ageing effect increases dramatically with humidity (Fig. 2b).

This humidity dependence leads to the intuitive idea that the ageing effect originates from the condensation of small liquid bridges between the beads. Liquid bridges induce a significant cohesion between the beads, which can increase the friction between different layers of the granular system and result in a higher value of θ_w . However, the physical justification for such

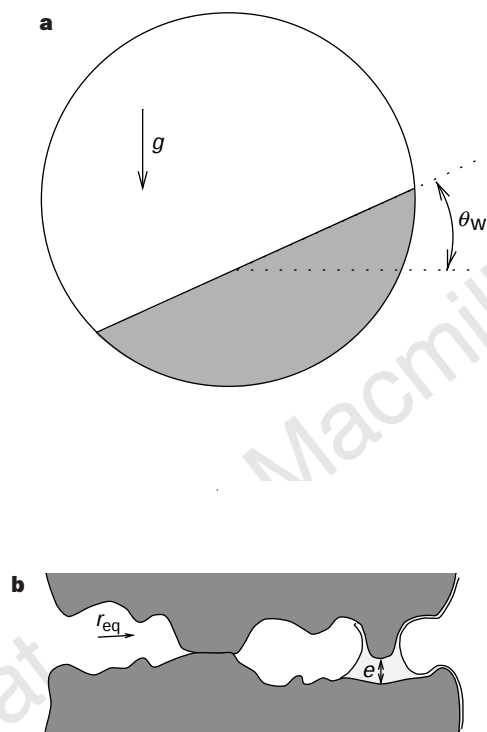


Figure 1 Experimental set-up. **a**, Principle of the experimental setup. Glass beads fill 25% of the volume of a cylinder (diameter 100 μm , length 13 mm) which can rotate around its axis. Two systems were studied; a 'polydisperse' system with $140 \mu\text{m} < d < 260 \mu\text{m}$, and a 'monodisperse' system with $200 \mu\text{m} < d < 250 \mu\text{m}$, where d is the diameter of the beads. The walls of the cylinder are glass, and the shape of the pile of beads is recorded with a video camera. Experiments are performed at room temperature, under controlled humidity (defined as the ratio between the vapour pressure and saturated pressure of water, P_v/P_{sat}). Before starting experiments, the system is prepared by rotating the drum for ~ 12 h. Then the ageing properties are investigated for various values of P_v/P_{sat} using the following protocol: first, the system is put in motion for a few turns (typically three); second, the end of this 'motion period' defines the origin of waiting time $t_w = 0$, after which the system is left at rest; third, after a given time (ranging from 10 to 10^4 s), a slow rotational motion (< 1 turn per min) is transmitted to the cylindrical drum. The angle θ_w at which the first avalanche takes place is measured from the slope of the beads just before the avalanche. The same procedure is repeated for different waiting times t_w and a curve of θ_w versus waiting time t_w is constructed. **b**, Schematic drawing of the contact at the nanometre scale between two micrometre-size asperities on the beads. Inside most of the contact region, the solid surfaces do not touch each other at the molecular scale, and capillary condensation occurs in the gap left between the surfaces.

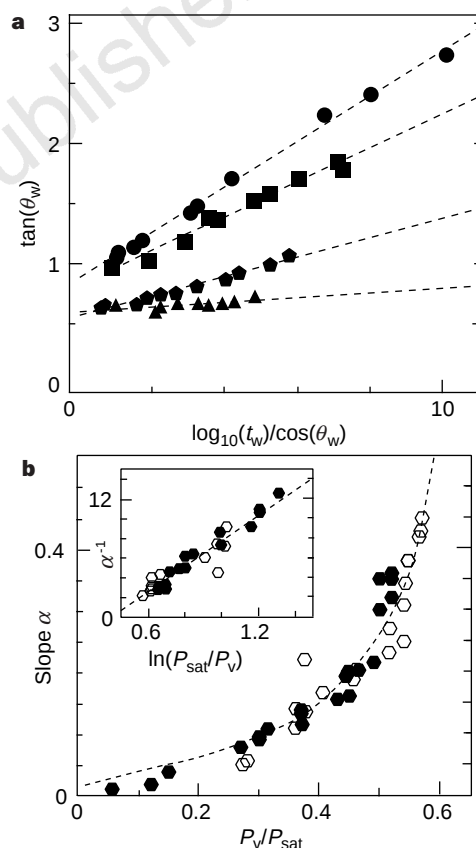


Figure 2 Dependence of ageing on humidity. **a**, Logarithmic ageing of the angle of first avalanche. The ageing property is analysed by plotting $\tan \theta_w(t_w)$ as a function of $\log_{10}(t_w)/\cos \theta_w(t_w)$, for different values of the water vapour pressure P_v (see equation (3)). The time is in units of seconds. From bottom to top, humidity is: 15% (triangles), 27% (pentagons), 36.1% (squares) and 45.5% (circles). In these measurements, times range typically over nearly four orders of magnitude. The dotted lines are least-square fits of the experimental data, whose slope is $\alpha(P_v)$. **b**, Variation of the slope $\alpha(P_v)$ characterizing the ageing behaviour of the first avalanche angle (see equation (3)) with humidity P_v/P_{sat} . Open symbols, the polydisperse system; filled symbols, the monodisperse system. The dashed line is the theoretical prediction $\alpha = \alpha_0/\ln(P_{\text{sat}}^*/P_v)$, where $\alpha_0 = 0.079$ and $P_{\text{sat}}^* = 0.68 P_{\text{sat}}$. The validity of the previous theoretical prediction is best confirmed by the measured linear dependence of $1/\alpha$ as a function of $\ln(P_{\text{sat}}/P_v)$ (see inset): in this plot, a linear least-squares fit thus provides unambiguously values of α_0 and P_{sat}^* . The lowering observed in the saturating pressure P_{sat}^* might be an effect of the long-range attractive forces exerted by the walls, and/or of dissolved species in the condensed water.

behaviour is not so obvious. We consider first the idealized situation of two smooth beads in contact. The capillary adhesion force exerted by a small liquid bridge connecting the beads is given by the product of the liquid surface tension γ and the radius of the beads (we note that if solid deformation occurs, the numerical prefactor changes slightly⁸): $F_{\text{adh}} \approx 2\pi\gamma R$, regardless of the size of the (small) bridge⁹. Therefore, the adhesion force has no humidity-dependence, and moreover would be able to stick all the beads together². This apparent difficulty disappears if one takes into account the roughness of the beads. The spatial extent of a liquid bridge is controlled by the radius of curvature r_{eq} of the liquid interface, which, at liquid–vapour equilibrium, is fixed by the Kelvin relation¹⁰

$$\frac{\gamma}{r_{\text{eq}}} = \rho_l k_B T \ln \frac{P_{\text{sat}}}{P_v} \equiv \rho_l \Delta\mu \quad (1)$$

where ρ_l is the density of the liquid and $\Delta\mu = k_B T \ln(P_{\text{sat}}/P_v)$ is the undersaturation of the chemical potential. Under ambient conditions, this yields a value of r_{eq} of nanometre dimensions. A crucial consequence is that liquid bridges are able to form only in nanometre-scale interstices. Because the beads are not smooth at the nanometre scale, the wetted region does not spread over the whole area it would occupy if the beads were smooth, and the cohesive force is reduced in proportion. We also note the slow evolution in time of the cohesive properties, measured through the time-dependence of θ_w . This indicates that the condensation of liquid bridges takes place over long timescales. However, surface-force experiments on capillary condensation between smooth surfaces separated by a gap show clearly that the condensation of a liquid bridge occurs from a metastable state; that is, a metastable vapour phase can stay for a long time (of the order of several minutes) between the surfaces, whereas the equilibrium state is a liquid bridge^{11,12}.

On the basis of these considerations, we propose a model for the ageing behaviour based on the hypothesis that capillary condensation arises through an activated process. This assumption is justified by the first-order character of capillary condensation. We consider two surfaces separated by a gap, and in contact with undersaturated vapour. A thin wetting film coats both surfaces. For gaps of less than a critical distance, of the order of the Kelvin radius r_{eq} , this state is metastable¹⁰: capillary condensation should occur. However, an energy barrier has to be overcome, as the coating films have to grow and coalesce in order to fill the gap between the surfaces. Along this path, the free energy of the system increases to a maximum as the films are about to merge¹¹. The energy barrier, ΔE , is therefore the free energy cost of condensing the corresponding water volume from the undersaturated vapour phase: $\Delta E \approx \Delta\mu \rho_l v_1$, where v_1 is the liquid volume needed to nucleate the liquid bridge. In the case of two rough beads, nucleation occurs preferentially between the asperities and the nucleation volume is $v_1 = a_0^2 e$, where e is the gap between the surfaces at the nucleating site and a_0^2 is a typical nucleation area (Fig. 1b). A more sophisticated model taking the long-range forces into account confirms this estimation.

Assuming an activation process, the time τ needed to condense a liquid bridge in an interstitial volume is $\tau \approx \tau_0 \exp(\Delta E/k_B T)$, with τ_0 being a microscopic time. Typically, τ_0 is of the order of the time needed to condense one liquid layer. Because both beads are rough, many liquid bridges can form in the macroscopic contact region. One expects, moreover, that the nucleating sites will exhibit a broad distribution of gaps e between solid surfaces, and that the activation times are accordingly widely distributed. After a given time t_w only the bridges with an activation time τ_{act} smaller than t_w have condensed. These were therefore formed at the nucleating sites with a gap e , verifying that $e < e_{\text{max}}(t_w) = (k_B T/\Delta\mu)(\rho_l a_0^2)^{-1} \ln(t_w/\tau_0)$. Once a liquid bridge has condensed, it fills locally the volume

surrounding the nucleating site, until the Kelvin equilibrium condition for the radius of curvature, equation (1), is met. Thus because of roughness, only a fraction $f(t_w)$ of the total wettable area is indeed wetted at a given time t_w . This fraction $f(t_w)$ is proportional to the number of activated bridges, yielding (to a first approximation) $f(t_w) \approx e_{\text{max}}(t_w)/\lambda$, where λ is the typical width of the distribution of distances between the surfaces. The capillary adhesion force is proportional to the wetted area and is thus reduced by the factor $f(t_w)$ from the perfectly smooth case ($F_{\text{adh}} \approx 2\pi\gamma R$), leading to

$$F_{\text{adh}}(t_w) \approx \gamma d \frac{1}{\ln(P_{\text{sat}}/P_v)} \ln\left(\frac{t_w}{\tau_0}\right) \quad (2)$$

where $d = 2\pi R/(\lambda \rho_l a_0^2)$ is a distance taking into account the geometrical characteristics of the contact. By reproducing Coulomb's argument for the stability of the surface layer in the presence of this additional adhesive force^{1,7}, we obtain the following equation for $\theta_w(t_w)$:

$$\tan \theta_w(t_w) \approx \tan \theta_0 + \frac{\alpha(P_v)}{\cos \theta_w(t_w)} \log\left(\frac{t_w}{t_0}\right) \quad (3)$$

where $\alpha(P_v) = \alpha_0/\ln(P_{\text{sat}}/P_v)$. Here α_0 is a dimensionless parameter taking into account the relative strength of the adhesion force and the weight of the material. Thus, by plotting $\tan \theta_w(t_w)$ as a function of $\log(t_w)/\cos \theta_w(t_w)$, one should obtain a straight line. This expectation is indeed borne out experimentally (Fig. 2a). Moreover, the increase in the slope, $\alpha(P_v)$, of this line with humidity is in good agreement with the theoretical prediction of equation (3) (Fig. 2b). Our model is thus able to reproduce both the waiting-time- and humidity-dependence of the measured ageing properties of a granular system.

Our results highlight the crucial role of humidity in the statics of granular systems. The similarity between ageing in granular media and ageing of the static friction coefficient in dry friction is reflected in the reported effect of humidity on the friction between rocks⁴; in that study, the 'standard' ageing properties of μ_s were similarly found to disappear at zero humidity. Similar behaviour has also been observed in indentation experiments¹³. The mechanism previously suggested to explain ageing behaviour in solid friction involves asperity creep¹⁴; our work suggests that humidity-induced capillary condensation may provide an alternative explanation for this effect. □

Received 26 January; accepted 2 September 1998.

1. Coulomb, C. A. Sur une application des règles de Maximis et Minimis à quelques Problèmes de Statique, relatifs à l'Architecture, in *Mémoires de Mathématiques et de Physique, Académie Royale des Sciences, Paris*, 343–382 (1773).
2. Duran, J. *Sables, Poudres et Grains* (Eyrolles Sciences, Paris, 1997).
3. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).
4. Dieterich, J. & Conrad, G. Effect of humidity on time- and velocity-dependent friction in rocks. *J. Geophys. Res.* **89**, 4196–4202 (1984).
5. Heslot, F., Baumberger, T., Perrin, B., Caroli, B. & Caroli, C. Creep, stick-slip, and dry-friction dynamics: Experiments and a heuristic model. *Phys. Rev. E* **49**, 4973–4988 (1994).
6. Hornbaker, D. J., Albert, R., Albert, I., Barabasi, A.-L. & Schiffer, P. What keeps sandcastles standing? *Nature* **387**, 765 (1997).
7. Halsey, T. C. & Levine, A. J. How sandcastles fall. *Phys. Rev. Lett.* **80**, 3141–3144 (1998).
8. Maugis, D. Adhesion of spheres: the JKR-DMT transition using a Dugdale model. *J. Colloid Interface Sci.* **150**, 243–269 (1992).
9. Israelachvili, J. N. *Intermolecular and Surface Forces* (Academic, London, 1985).
10. Evans, R. in *Liquids at Interfaces* (eds Charvolin, J., Joanny, J. F. & Zinn-Justin, J.) 3–98 (Elsevier Science, New York, 1989).
11. Crassous, J., Charlaix, E. & Loubet, J.-L. Capillary condensation between high-energy surfaces. An experimental study with a surface force apparatus. *Europhys. Lett.* **28**, 37–42 (1994).
12. Iwamatsu, I. & Horii, K. Capillary condensation and adhesion of two wetted surfaces. *J. Colloid Interface Sci.* **182**, 400–406 (1996).
13. Westbrook, J. H. & Jorgensen, P. J. Effect of water desorption on indentation microhardness anisotropy in minerals. *Am. Mineral.* **53**, 1899–1914 (1968).
14. Estrin, Y. & Brechet, Y. On a model of frictional sliding. *Pure Appl. Geophys.* **147**, 745–762 (1996).

Acknowledgements. We thank J.-M. Georges and M.L. Bocquet for discussions. L.B. thanks J. Duran for providing ref. 1.

Correspondence and requests for materials should be addressed to L.B. (e-mail: lbocquet@physique.ens-lyon.fr).

Coherent moving states in highway traffic

Dirk Helbing* & Bernardo A. Huberman†

* II Institute of Theoretical Physics, University of Stuttgart, Pfaffenwaldring 57/III, 70550 Stuttgart, Germany

† Xerox PARC, 3333 Coyote Hill Road, Palo Alto, California 94304, USA

Advances in multiagent simulation techniques^{1–3} have made possible the study of realistic highway traffic patterns and have allowed theories^{3–6} based on driver behaviour to be tested. Such simulations display various empirical features of traffic flows⁷, and are used to design traffic controls that maximize the throughput of vehicles on busy highways. In addition to its intrinsic economic value⁸, vehicular traffic is of interest because it may be relevant to social phenomena in which diverse individuals compete with each other under certain constraints^{9,10}. Here we report simulations of heterogeneous traffic which demonstrate that cooperative, coherent states can arise from competitive interactions between vehicles. As the density of vehicles increases, their interactions cause a transition into a highly correlated state in which all vehicles move with approximately the same speed, analogous to the motion of a solid block. This state is safer because it has a reduced lane-changing rate, and the traffic flow is high and stable. The coherent state disappears when the vehicle density exceeds a critical value. We observe the effect also in real Dutch traffic data.

In many social situations, decisions made by individuals lead to external effects that may be very regular, even without a global coordinator. In traffic, these decisions concern when to accelerate or brake, to overtake or to enter a busy multi-lane road^{11–14}, while trying to get ahead as fast as possible, but safely, under the constraints imposed by physical limitations and traffic rules. At times these behaviours give rise to very regular traffic patterns, as exemplified by the universal characteristics of moving traffic-jam fronts¹⁵ or synchronized congested traffic^{16,17}. These phenomena are in contrast with usual social dilemmas, where cooperation in order to achieve a desirable collective behaviour hinges on having small groups or long time horizons^{9,18}.

Here we describe a type of collective behaviour that we discovered when studying the dynamics of a diverse set of vehicles, such as cars and lorries, travelling along a two-lane highway with different velocities. As the density of vehicles in the road increases, there is a transition into a highly coherent state characterized by all vehicles having the same average velocity and a very small dispersion around this value. The transition to this behaviour becomes apparent when looking at the travel-time distributions of cars and lorries, comprising the overall dynamics on a stretch of highway (Fig. 1). These travel times were obtained by running computer simulations using a discretized follow-the-leader algorithm^{19,20}, which distinguishes two neighbouring lanes i of a unidirectional highway. Both are subdivided into sites $z \in \{1, 2, \dots, L\}$ of equal length $\Delta x = 2.5$ m. Each site is either empty or occupied, the latter case representing the back of a vehicle of type a (for example, a car or lorry) with velocity $v = u\Delta x/\Delta t$. Here $u \in \{0, 1, \dots, u_a^{\max}\}$ is the number of sites that the vehicle moves per update step $\Delta t = 1$ s. Cars and lorries are characterized by different 'optimal' or 'desired' (that is, maximally safe) velocities $U_a(d_+)$ with which the vehicles would like to drive at a distance d_+ to the vehicle in front (see symbols in Fig. 2). Their lengths l_a correspond to the maximum distances satisfying $U_a(l_a) = 0$. At times $T \in \{1, 2, \dots\}$, that is, every time step Δt , the positions $z(T)$, velocities $u(T)$ and lanes $i(T)$ of all vehicles are updated in parallel. We have ruled out synchronization artefacts²¹ by this update method, which is appropriate for flow simulations¹.

Denoting the position, velocity and distance of the respective

leading vehicle (+) or following vehicle (–) on lane $i(T)$ by $z_{\pm}$, $u_{\pm}$ and $d_{\pm} = |z_{\pm} - z|$, in the adjacent lane by $z'_{\pm}$, $u'_{\pm}$ and $d'_{\pm} = |z'_{\pm} - z|$, the successive update steps are:

1. Determine the potential velocities $u(T+1)$ and $u'(T+1)$ on the present and the adjacent lane according to the acceleration law²²

$$u^{(i)}(T+1) = [\lambda U_a(d_+^{(i)}(T)) + (1-\lambda)u(T)]$$

where the floor function $[x]$ is defined by the largest integer $l \leq x$. This describes the typical follow-the-leader behaviour of driver-vehicle units. Delayed by the reaction time Δt , they tend to move with their desired velocity U_a , but the adaptation takes a certain time $\tau = \lambda\Delta t$ because of the vehicle's inertia.

2. Change lane in accordance with, for simplicity, symmetrical ('American') rules, if the following incentive and safety criteria¹⁴ are fulfilled: check if the distance $d'_-(T)$ to the following vehicle in the neighbouring lane is greater than the distance $u'_-(T+1)$ that this vehicle is expected to move within the reaction time Δt (safety criterion 1). If so, the difference $D = d'_-(T) - u'_-(T+1) > 0$ defines the backward surplus gap. Next, look if you could go faster in the adjacent lane (incentive criterion). Finally, make sure that the relative velocity $[u'_-(T+1) - u'(T+1)]$ would not be

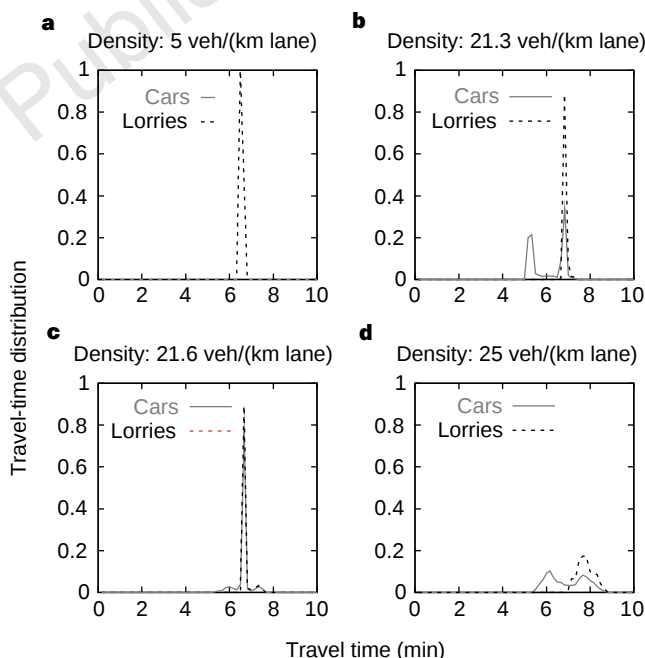


Figure 1 Simulated travel-time distributions for a circular one-way two-lane highway of 10 km length. The chosen model parameters $\lambda = 0.77$, $p = 0.05$ and $m = 2$ (see text), together with the desired velocity functions shown in Fig. 2, yield a good representation of the dynamics on fast lanes of Dutch highways²². We considered a scenario of ~2.5% randomly selected, identical lorries of length 7.5 m with a maximum velocity of ~90 km h⁻¹ moving among 97.5% identical cars with 5 m length and a maximum velocity of ~125 km h⁻¹. **a**, At small densities, the travel-time distribution has two narrow peaks at the maximum velocities of cars and lorries, as cars can overtake lorries without prior slow-downs. **b**, With increasing but still moderate density, the travel times of lorries remain unchanged, as their slow speed implies, on average, large distances to the next vehicles ahead. In contrast, the average travel time of cars grows. As a lack of sufficiently large gaps may prevent immediate lane changing, some cars will have to temporarily slow down to the lorries' speed. The resulting higher relative velocity to the vehicles in the adjacent lane makes overtaking more difficult, so they may get 'trapped' behind lorries for a long time. Therefore, the travel-time distribution develops a second peak around the lorry peak. **c**, At a certain density, the peak of unobstructed cars disappears, and the travel-time distributions of cars and lorries become almost identical, with a small dispersion. **d**, If the density is further increased, this highly correlated state of motion breaks down, and a broad distribution of travel times results.

larger than $D + m$ (safety criterion 2), where the magnitude of the parameter $m \geq 0$ is a measure of how aggressive drivers are in overtaking (by possibly enforcing deceleration manoeuvres).

3. If, in the updated lane $i(T + 1)$, the corresponding potential velocity $u(T + 1)$ or $u'(T + 1)$ is positive, diminish it by 1 with probability p in order to account for delayed adaptation and the variation of vehicle velocities.

4. Update the vehicle position according to the equation of motion $z(T + 1) = z(T) + u(T + 1)$.

Despite its simplifications, this model is in good agreement with the empirical known features of traffic flows, and it can be well calibrated²²: λ and $U_a(d_+)$ determine the approximate velocity–density relation and the instability region. The typical outflow from traffic jams and their characteristic dissolution velocity¹⁵ can be enforced by Δt and Δx . The average distance between successive traffic jams increases with smaller p . The parameter m allows us to calibrate the lane-changing rates. Our simulations started with uniform dis-

tances among the vehicles and their associated desired velocities. The lorries were randomly selected. As our evaluations started after a transient period of one hour and extended over another four hours, the results are largely independent of the initial conditions.

Investigating the density-dependent average velocities of cars and lorries yields further insight into the solid-like state (Fig. 2). For small p one finds that, at a certain ‘critical’ density, the average speed of cars decays significantly towards the speed of the lorries, which is still close to their maximum velocity. At this density, the highway space is almost used up by the safe vehicle distances, so that sufficiently large gaps for lane-changing can only occur for strongly varying vehicle velocities (like for large p). However, because the speeds of cars and lorries are almost identical in the solid-like state, the lane-changing rate drops by almost one order of magnitude (Fig. 3). Consequently, without opportunities for overtaking, all vehicles have to move coherently at the speed of the lorries, which closes the feedback loop that causes the transition. The solid-like flow does

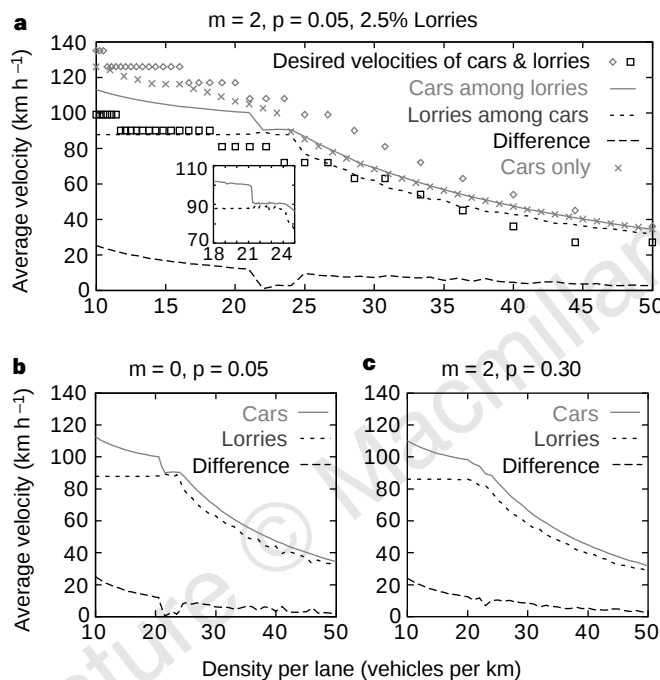


Figure 2 Numerically determined average velocities of cars (solid lines) and lorries (short-dashed lines) as a function of the overall vehicle density. **a**, Two transitions are observed. 1. Up to a density of 24 vehicles per km, the average velocity of lorries remains unchanged. This is analogous to the dynamics of mixed one-lane traffic²⁷: because of their slower speed, lorries experience smaller local densities as long as the average velocity of pure car traffic (crosses) stays above the maximum velocity of lorries, so that the road is not completely used by the vehicular space requirements at this speed. Then, the average lorry speed falls significantly with growing density to maintain safe vehicle distances. This causes an instability of traffic flow^{6,20}, resulting in a formation of stop-and-go traffic. 2. At a density of ~ 21.5 vehicles per km, the average velocity of cars drops almost to the average velocity of lorries, which leads to a distinct minimum in the difference of both curves (long-dashed line). This transition seems to be quite sharp, as illustrated by the inset at greater density resolution. Between 21.5 and 24 vehicles per km, the vehicles move like a solid block. Obviously, this is not enforced by the difference between the assumed dependencies of the desired velocities of cars (diamonds) and lorries (squares) on the local density ahead of them, defined by the inverse vehicle distance. The parameter values were chosen as in Fig. 1, but the observed effects are not very sensitive to the particular choice. Different values of m give the same qualitative results (**b**). A higher proportion of lorries leads to an earlier transition to solid-like behaviour. Increasing the fluctuation parameter p causes a smoother, but still visible, transition, until it disappears for $p \approx 0.3$ (**c**). A similar effect occurs for the width of velocity or parameter distributions characterizing cars and lorries.

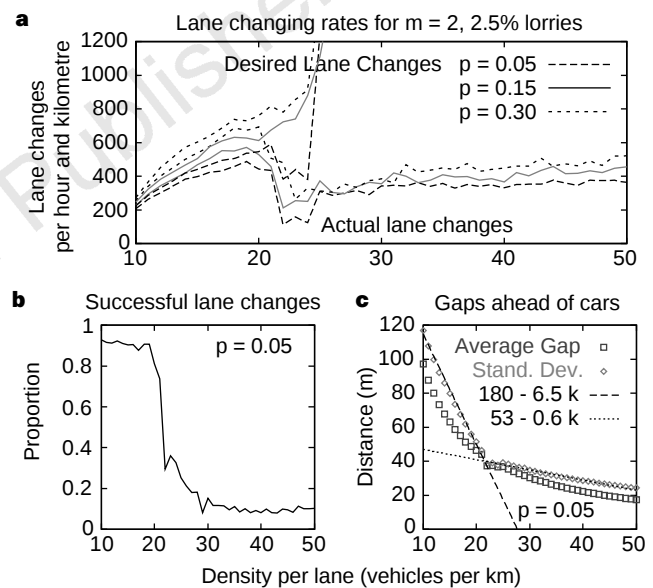


Figure 3 Dependence of desired and actual lane-changing rates and associated quantities on the overall vehicle density. **a**, The simulated actual lane-changing rates break down in the density range of coherent motion. Here, the parameters are the same as in Figs 1 and 2. For stronger fluctuations p , the minimum is less pronounced, but still noticeable up to $p < 0.3$. Smaller values of m reduce the number of lane changes, but do not prevent the solid-like state. We have checked whether the lane-changing rate breaks down because, with identical velocities of cars and lorries in both lanes, there may be no advantage to lane changing. However, if the rates of desired lane changes (according to the incentive criterion) are reduced at all, they still keep a high level. **b**, The transition point around 21.5 vehicles per km is characterized by a rapid decay of the proportion of successful lane changes (that is the quotient between actual and desired lane changes). **c**, With growing density, not only the average of the gaps in front of cars decreases (squares), but also their standard deviation (diamonds). Opportunities for lane-changing are rapidly diminished when gaps of about twice the safe vehicle distance required for lane-changing cease to exist. This relates to a significantly smaller slope of the gaps' standard deviation after the solid-block transition (broken lines; k denotes the density per lane). The breakdown of the lane-changing rate seems to imply a decoupling of the lanes, that is, an effective one-lane behaviour. However, this is already the result of a self-organization process based on two-lane interactions, as any significant perturbation of the solid-block state (such as different densities in neighbouring lanes or velocity variations) will cause frequent lane changes. By filling large gaps, the gap distribution is considerably modified (also compared to mixed one-lane traffic²⁷). This will eventually reduce possibilities for lane changes, so that the solid-like state is restored.

not change by adding vehicles until the whole highway is saturated by the vehicular space requirements at the speed of the lorries. Then, the vehicle speeds decay significantly to maintain safe vehicle distances. The onset of stop-and-go traffic at this density produces gaps that vary by a large amount, so that overtaking is again possible and the coherent state is destroyed. For large p , we do not have a breakdown of the lane-changing rate at the critical density and, hence, no coherent state. Nevertheless, lane-changing cars begin to interfere with the lorries, so that the average velocity of lorries starts to decrease with growing density before the average car velocity comes particularly close to it (Fig. 2c).

As shown in Fig. 4, our prediction of the transition into a coherent state is supported by empirical data obtained from highway traffic in the Netherlands. Clearly, the difference between the average velocities of cars and lorries shows the predicted minimum at a density around 25 vehicles per kilometre and lane, where the average car velocity approaches the constant velocity of the lorries (Fig. 4a, b). The fact that the empirically observed minimum is less distinct than in Fig. 2a can be reproduced by higher values of the

fluctuation parameter ($p \approx 0.15$) and points to a noisy transition. This interpretation is also supported by the relative fluctuation of vehicle speeds (Fig. 4c), which shows a minimum at the same density, while remaining finite. A similar result is obtained for the interaction rates of vehicles (Fig. 4d).

We have presented here a new effect in highway traffic that consists of the formation of coherent motion out of a disorganized vehicle flow by competitive interactions. The predicted solid-like state is supported by real highway data, and our interpretation of the effect suggests that it is largely independent of the chosen driver-vehicle model (although the transition may be less sharp in a continuous model). It would be interesting to see if the spontaneous appearance of coherent states is also found in other social or biological systems, such as pedestrian crowds²³, cell colonies²⁴ or animal swarms²⁵.

We note that the coherent state of vehicle motion considerably reduces the main sources of highway accidents: differences in vehicle speeds and lane-changes²⁶. It is also associated with maximum throughput in the highway, and is located just before the transition to unstable traffic flow. Thus, at a practical level it is desirable to implement traffic rules and design highway controls that will lead to traffic's moving like a solid block. Close to the transition point, the formation of this coherent state could be supported by traffic-dependent lane-changing restrictions and variable speed limits or by automatic vehicle control systems. Compared with 'American' (symmetric) lane-changing rules, 'European' rules seem to be less efficient: an asymmetric lane usage, where lorries mainly keep in one lane and overtaking is carried out in the other lane(s), motivates car drivers to avoid the lorry lane, so that the effective highway capacity is reduced by up to 25%. □

Received 30 April; accepted 16 September 1998.

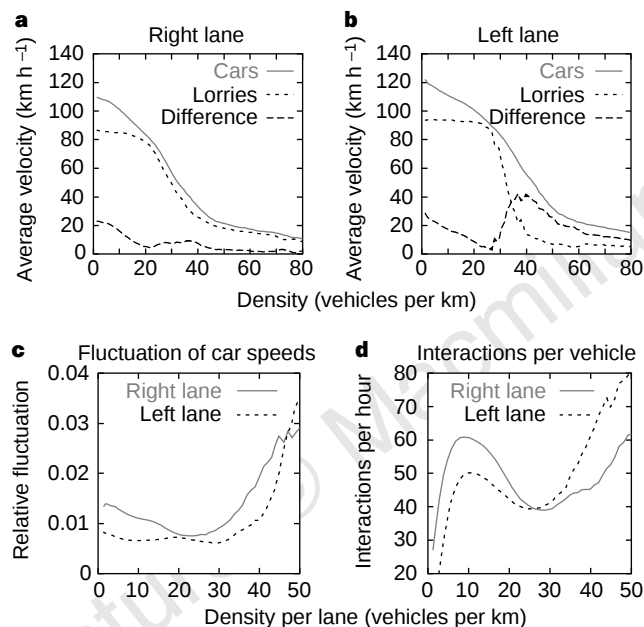


Figure 4 Mean values of one-minute averages that were determined from single-vehicle data on mixed traffic on the Dutch two-lane highway A9 on 14 subsequent days. Panels **a** and **b** depict the density-dependent average velocities of cars (solid lines) and lorries (short-dashed lines) in the right and left lane, respectively. Their differences (long-dashed lines) show the predicted minimum at densities of ~ 25 vehicles per km, up to which the velocity of lorries is almost constant. **c**, The relative fluctuations of car speeds (defined as velocity variance divided by the square of average velocity) display minima at the same densities, which points to a more coherent motion in the car fraction. **d**, The interaction rates per vehicle also show a minimum around 25 vehicles per km. This corresponds to a decreased relative velocity among successive vehicles. The interaction rate is defined by the average of $\min(\Delta v/d, 0)$, where Δv denotes the velocity difference to the vehicle in front. We point out that the above data support the predictions of our model despite its simplifications. In particular, this concerns the asymmetry of European lane-changing rules, which imply that overtaking is allowed only in the left lane, but vehicles should switch back to the right lane as soon as possible. At least at velocities of ≈ 80 km h⁻¹ vehicles also pass in the right lane with small relative velocities to vehicles in the left lane. Nevertheless, the rate of lane changes from and to each lane is, on average, the same. One result of the mentioned asymmetry, however, is a smaller fraction of lorries in the left lane, which diminishes the related minimum in **c**. Another consequence is the tendency to have a higher average car velocity in the left lane, so that in **b**, compared to **a**, the minimum in the difference between the average speeds of cars and lorries is more pronounced.

- Schreckenberg, M., Schadschneider, A., Nagel, K. & Ito, N. Discrete stochastic models for traffic flow. *Phys. Rev. E* **51**, 2939–2949 (1995).
- Faieta, B. & Huberman, B. A. *Firefly: A Synchronization Strategy for Urban Traffic Control* (Internal Rep. No. SSL-42, Xerox PARC, Palo Alto, CA, 1993).
- Schreckenberg, M. & Wolf, D. E. *Traffic and Granular Flow '97* (Springer, Singapore, 1998).
- Prigogine, I. & Herman, R. *Kinetic Theory of Vehicular Traffic* (Elsevier, New York, 1971).
- Leutzbach, W. *Introduction to the Theory of Traffic Flow* (Springer, Berlin, 1988).
- Helbing, D. *Verkehrsdynamik* (Springer, Berlin, 1997).
- May, A. D. *Traffic Flow Fundamentals* (Prentice Hall, Englewood Cliffs, NJ, 1990).
- Small, K. A. *Urban Transportation Economics* (Harwood Academic, London, 1992).
- Glaube, N. S. & Huberman, B. A. The dynamics of social dilemmas. *Sci. Am.* **270**, 76–81 (1994).
- Helbing, D. *Quantitative Sociodynamics* (Kluwer Academic, Dordrecht, 1995).
- Daganzo, C. F. Probabilistic structure of two-lane road traffic. *Transportation Res.* **9**, 339–346 (1975).
- Hall, F. L. & Lam, T. N. The characteristics of congested flow on a freeway across lanes, space, and time. *Transportation Res. A* **22**, 45–56 (1988).
- Rickert, M., Nagel, K., Schreckenberg, M. & Latour, A. Two lane traffic simulations using cellular automata. *Physica A* **231**, 534–550 (1996).
- Nagel, K., Wolf, D. E., Wagner, P. & Simon, P. Two-lane traffic rules for cellular automata: A systematic approach. *Phys. Rev. E* **58**, 1425–1437 (1998).
- Kerner, B. S. & Rehborn, H. Experimental features and characteristics of traffic jams. *Phys. Rev. E* **53**, R1297–R1300 (1996).
- Kerner, B. S. & Rehborn, H. Experimental properties of phase transitions in traffic flow. *Phys. Rev. Lett.* **79**, 4030–4033 (1997).
- Helbing, D. & Treiber, M. Gas-kinetic-based traffic model explaining observed hysteretic phase transition. *Phys. Rev. Lett.* **81**, 3042–3045 (1998).
- Axelrod, R. & Dion, D. The further evolution of cooperation. *Science* **242**, 1385–1390 (1988).
- Gazis, D. C., Herman, R. & Rothery, R. W. Nonlinear follow the leader models of traffic flow. *Operations Res.* **9**, 545–567 (1961).
- Bando, M. et al. Phenomenological study of dynamical model of traffic flow. *J. Phys. I France* **5**, 1389–1399 (1995).
- Huberman, B. A. & Glaube, N. S. Evolutionary games and computer simulations. *Proc. Natl Acad. Sci. USA* **90**, 7716–7718 (1993).
- Helbing, D. & Schreckenberg, M. Cellular automata simulating experimental properties of traffic flows. *Phys. Rev.* (submitted).
- Helbing, D., Keltsch, J. & Molnár, P. Modelling the evolution of human trail systems. *Nature* **388**, 47–50 (1997).
- Ben-Jacob, E. et al. Generic modelling of cooperative growth patterns in bacterial colonies. *Nature* **368**, 46–49 (1994).
- Vicsek, T. et al. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**, 1226–1229 (1995).
- Lave, C. A. Speeding, coordination, and the 55 MPH limit. *Am. Econ. Rev.* **75**, 1159–1164 (1985).
- Krug, J. & Ferrari, P. A. Phase transitions in driven diffusive systems with random rates. *J. Phys. A: Math. Gen.* **29**, L465–L471 (1996).

Acknowledgements. D.H. thanks the DFG for support by a Heisenberg scholarship, and H. Taale and the Dutch Ministry of Transport, Public Works and Water Management for supplying the highway data, which were evaluated by V. Shvetsov.

Correspondence and requests for materials should be addressed to D.H. (e-mail: helbing@theo2.physik.uni-stuttgart.de).

Elasticity and rheology of iron above 220 GPa and the nature of the Earth's inner core

Ho-kwang Mao*, Jinfu Shu*, Guoyin Shen†, Russell J. Hemley*, Baosheng Li‡ & Anil K. Singh§

* Geophysical Laboratory and Center for High Pressure Research, Carnegie Institution of Washington, 5251 Broad Branch Road, NW, Washington DC 20015, USA

† Consortium for Advanced Radiation Sources, University of Chicago, Chicago, Illinois 60637, USA

‡ Center for High Pressure Research, Mineral Physics Institute, State University at Stony Brook, Stony Brook, New York 11794-2100, USA

§ Materials Science Division, National Aerospace Laboratories, Bangalore 560017, India

Recent numerical-modelling and seismological results have raised new questions about the dynamics^{1,2} and magnetism^{3,4} of the Earth's core. Knowledge of the elasticity and texture of iron^{5,6} at core pressures is crucial for understanding the seismological observations, such as the low attenuation of seismic waves, the low shear-wave velocity^{7,8} and the anisotropy of compressional-wave velocity^{9–11}. The density and bulk modulus of hexagonal-close-packed iron have been previously measured to core pressures by static¹² and dynamic^{13,14} methods. Here we study, using radial X-ray diffraction¹⁵ and ultrasonic techniques¹⁶, the shear modulus, single-crystal elasticity tensor, aggregate compressional- and shear-wave velocities, and orientation dependence of these velocities in iron. The inner core shear-wave velocity is lower than the aggregate shear-wave velocity of iron, suggesting the presence of low-velocity components or anelastic effects in the core. Observation of a strong lattice strain anisotropy in iron samples indicates a large (~24%) compressional-wave anisotropy under the isostress assumption, and therefore a perfect alignment of crystals⁶ would not be needed to explain the seismic observations. Alternatively the strain anisotropy may indicate stress variation due to preferred slip systems.

Two radial X-ray diffraction (RXD) experiments (runs 1 and 2; see Fig. 1) were performed with diamond cells. Non-hydrostatic stress, a condition essential to the technique, was deliberately produced in the specimen by not adding a pressure medium. The stress state of the specimen compressed between two anvils is a superposition of the hydrostatic pressure

$$\sigma_p = (\sigma_3 + 2\sigma_1)/3 \quad (1)$$

and the differential stress components

$$t = \sigma_3 - \sigma_1 = 1.5(\sigma_3 - \sigma_p) \quad (2)$$

where σ_3 and σ_1 are axial and radial stresses, respectively¹⁷. A microfocus (4–10- μm diameter) polychromatic X-ray beam, which passed through the Be gasket in the radial direction,

probed the lattice strain of the sample as a function of the angle (ψ) to the diamond-cell axis¹⁵. At 21 pressures between 16 and 211 GPa, energy dispersive X-ray diffraction (EDXD) patterns containing (hkl notation) 100, 002, 101, 102, 110, 103, 112, 201 diffraction lines of hexagonal close packed (h.c.p.) iron, 111, 200, 220, 311, 222, 400, 331, 420, 422, 511 of gold, or 110, 200, 211, 220, 310, 222, 321, 400 of tungsten, were collected at 10° steps of ψ from 0° to 90° . The d -spacing varies linearly with $\cos^2 \psi$:

$$d(hkl) = d_p(hkl)[1 + (1 - 3\cos^2 \psi)Q(hkl)] \quad (3)$$

where the intercept $d_p(hkl)$ denotes the d -spacing under σ_p and the slope $Q(hkl)$ is the lattice strain under the uniaxial stress condition^{18,19}.

In run 1, a separate gold layer was used as a standard for determination of t and shear modulus G of h.c.p. Fe. The axial stress is continuous across the interface between the gold and iron layers ($\sigma_{3\text{Au}} = \sigma_{3\text{Fe}}$; subscripts denote the Au or Fe layer), that is

$$t_{\text{Fe}} = 1.5(\sigma_3 - \sigma_{\text{pFe}}) = 1.5(\sigma_{\text{pAu}} - \sigma_{\text{pFe}}) + t_{\text{Au}} \quad (4)$$

Now, t is related to G by

$$t = 6G\langle Q \rangle \quad (5)$$

where $\langle Q \rangle$ denotes the average value of measured Q for all (hkl) (ref. 15). The hydrostatic stress components, σ_{pAu} and σ_{pFe} , were determined from the observed $d_p(hkl)$ (equation (3)) and the equations of state of Au and Fe (refs 20, 21); G_{Au} was extrapolated from low-pressure data^{22,23}. The aggregate compressional-wave speed (v_p) and shear-wave speed (v_s) of h.c.p. Fe are calculated from the bulk modulus K_{Fe} and G_{Fe} . In addition (run 3), the aggregate ultrasonic

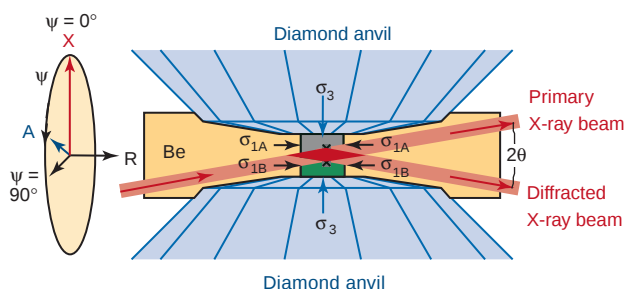


Figure 1 Lattice strains of diamond-cell samples under uniaxial stress (σ_1 and σ_3) are obtained with radial X-ray diffraction (RXD) through a beryllium gasket. The subscripts A and B denote the sample and stress standard. The vertical scale of the sample region is expanded. In run 1, the specimen, which consisted of a 15- μm (thickness) layer of an iron sample and a 3- μm layer of a gold standard, was compressed in a 40- μm (diameter) hole in a beryllium gasket between flat diamond anvils of 400- μm diameter. In run 2, the specimen consisted of a 5- μm iron and a 2- μm tungsten layer in a 15- μm hole in a Be gasket between bevelled diamond anvils (500- μm outer diameter, 90- μm inner diameter, 9.5° bevel angle). The diamond cell was mounted on a rotation stage with the diamond anvil axis (A) perpendicular to the rotation axis (R) which bisected the 2θ angle between incident and diffracted X-ray beams. The angle (ψ) between A and the diffraction vector (X) was varied as the diamond cell rotated around the R axis.

Table 1 Elasticity of h.c.p. Fe at 298 K and high pressures

P (GPa)	Density (g cm ⁻³)	C ₁₁ (GPa)	C ₁₂ (GPa)	C ₃₃ (GPa)	C ₁₃ (GPa)	C ₄₄ (GPa)	K (GPa)	G (GPa)	v _p (km s ⁻¹)	v _s (km s ⁻¹)	Run no.
16.5	9.00*						297	108	6.95	3.47	3
39	9.67*	504	244	493	254	271	351*	134	6.84	3.76	1
39	10.09	747	301	802	297	215	455	224	7.86	4.72	†
211	12.61‡	1303	637	1302	637	960	1071‡	396	10.42	5.61	2
211	12.80	1697	809	1799	757	421	1085	445	10.54	5.90	†

Runs 1, 2 and 3 are results of this study. They are compared with values from first-principles calculations⁶. (The method and results of ref. 24 are similar to those of ref. 6.) Uncertainties for K and v_p , 10% (runs 1 and 2), 2% (run 3); for C_{ij} , G and v_s , 20% (runs 1 and 2), 2% (run 3).

* Ref. 21.

† Ref. 6.

‡ Ref. 12.

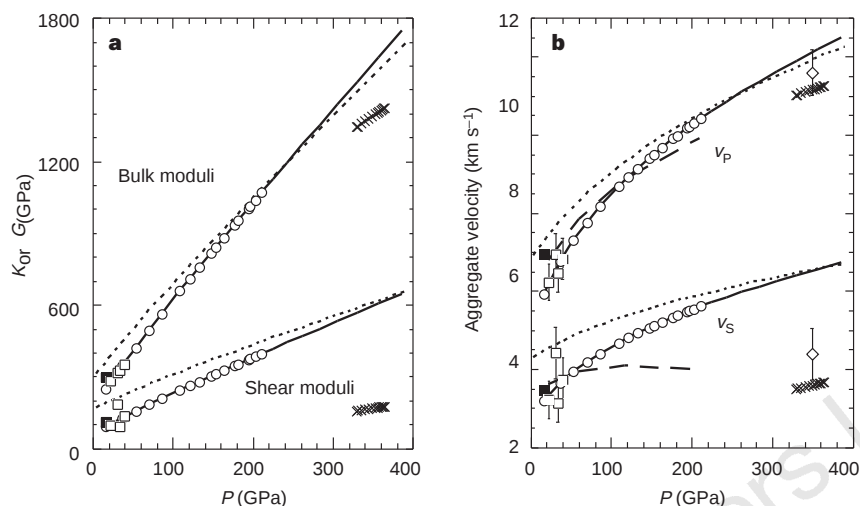


Figure 2 Comparison of our results with theoretical and shock-wave studies, and with seismic observations²⁵ in the inner core (crosses). Parameters compared are shear moduli and bulk moduli (**a**), and aggregate v_s and v_p (**b**) of h.c.p. Fe. Filled squares, ultrasonic measurements with multianvil apparatus in run 3; open squares, RXD measurements in run 1; solid curves, extrapolation of X-ray data in run 2 based on $K/G = 2.7$; short-dashed curves, first-principles calculations⁶ for 298 K isotherm; long-dashed curves, shock-wave Hugoniot at high temperatures^{13,14}. The solid curve was also used to calculate G and t for run 2 (open circles) for elasticity tensor estimations to 211 GPa. At 200 GPa, the differences between the Hugoniot values at 4407 K and the present 298 K values yield

temperature derivatives; for example, $-dv_s/dT = 3.7 \times 10^{-4} \text{ km s}^{-1} \text{ K}^{-1}$ and $-dv_p/dT = 0.9 \times 10^{-4} \text{ km s}^{-1} \text{ K}^{-1}$. In comparison with other estimations based on theory of thermal and elastic properties³⁰, this value of $-dv_s/dT$ is high, and $-dv_p/dT$ is low, possibly owing to the combined uncertainty in extrapolation of two data sets from entirely different studies (shock-wave and static compressions). If these derivatives are used for temperature correction of the solid curve to inner-core conditions, the resulting v_s and v_p of h.c.p. Fe at 6,000 K and 350 GPa (open diamonds in **b**) represent upper and lower bounds, respectively; that is, the differences between open diamonds and crosses is significant for v_s , but not for v_p .

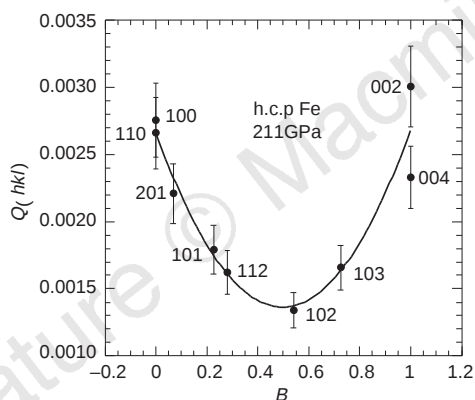


Figure 3 The measured $Q(hkl)$ of h.c.p. Fe at 211 GPa. This parameter follows a quadratic relation

$$Q(hkl) = m_0 + m_1 B + m_2 B^2 \quad (6)$$

where $B = 3a^2/l^2 [4c^2(h^2 + hk + l^2) + 3a^2/l^2]$. Under isostress assumption, the parameters m_0 , m_1 and m_2 provide three independent linear equations for determination of elasticity tensors^{15,27}.

v_p and v_s of h.c.p. Fe are measured directly at 16.5 GPa in a multianvil press¹⁶. All experiments were performed at 298 K.

The results from RXD (zero-frequency) and multianvil (ultrasonic-frequency) measurements are in good agreement, bracketing a wide frequency range including seismic waves. The observed K/G ratio of h.c.p. Fe is $2.7(\pm 0.7)$ from the RXD at 20–39 GPa, and is $2.68(\pm 0.1)$ from the ultrasonic measurement at 16.5 GPa. We extrapolate G_{Fe} , t_{Fe} , v_p and v_s to higher pressures based on a constant K/G of 2.7, and

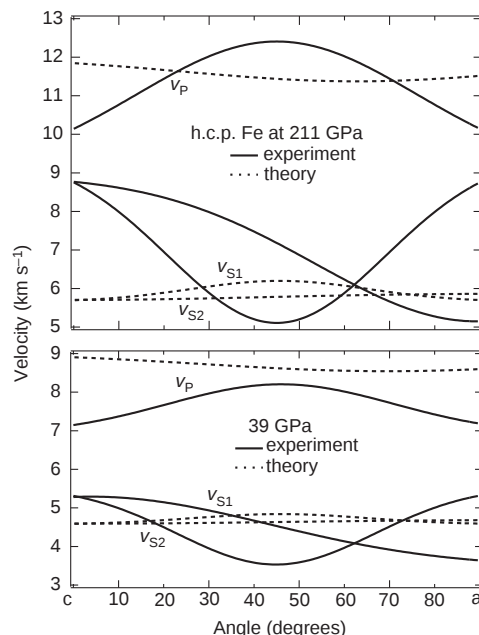


Figure 4 Comparison of seismic-wave velocities from this work with values calculated from theory. The velocities (v_p , v_{s1} and v_{s2}) of single-crystal h.c.p. Fe depend on the wave propagation direction relative to the c axis of the crystal. Solid curves, this study with isostress assumption; dashed curves, first-principles theory⁶.

compare the results with theoretical, shock-wave^{13,14} and seismic values (Fig. 2). The K_{Fe} , G_{Fe} , v_p and v_s predicted by first-principles theory^{6,24} are higher than the present values at low pressures, but the difference diminishes at higher pressures. Temperature derivatives of v_p and v_s are estimated from the differences between the Hugoniot data at high temperatures and the present static data at 298 K, and are used for estimation of v_p and v_s of h.c.p. Fe under the inner-core conditions²⁵ (Fig. 2b). The seismic v_s of the inner core is lower than the

corresponding h.c.p. Fe value. If the softening of v_s represents possible Born–Durand premelting effects²⁶ or partial melting, the observed shear values would tightly constrain the inner-core temperature near melting. The near-melting softening would also attenuate seismic waves. Alternatively, the observations may indicate the presence of additional low- v_s phases in the inner core.

The RXD measurements also provide single-crystal elasticity information^{15,27}. Figure 3 shows that the $Q(hkl)$ reaches a maximum at a (100, 110) and c (002) axes that is more than double the minimum of the $Q(hkl)$ at the diagonals (112, 102, 103). This trend persists over the entire pressure range studied. The strong (hkl) dependence of lattice strain reflects a strong elasticity anisotropy or an (hkl) dependence of stress^{18,19}, which has little effect on the aforementioned estimates of averaged aggregate properties^{15,27}, but which gives information on single-crystal properties. Elasticity tensors and the orientation dependence of v_p , v_{s1} and v_{s2} (subscripts 1 and 2 denote two polarization directions) are calculated from parameters in equation (6) (see Fig. 3 legend) at the isostress limit^{15,28}. Examples for h.c.p. Fe at 39 and 211 GPa are shown in Table 1 and Figs 2 and 4, and compared with values from first-principles theoretical predictions⁶. At 211 GPa (Fig. 3b), theories predicted a small v_p anisotropy (4% faster in the c than in the a direction) which requires a perfect alignment of h.c.p. Fe crystals (or a giant single crystal⁶) to account for the 4% inner-core v_p anisotropy. The present results obtained under the isostress assumption show a large v_p anisotropy (24% faster at 45° from c than along either the a or the c axis), which relieves the ‘perfect alignment’ textural constraint. Partial alignment of h.c.p. Fe crystals would be sufficient for the magnitude of inner-core anisotropy.

The isostress assumption for interpreting the data observed with our method has been confirmed experimentally for cubic phases of FeO and Fe (refs 15, 27). Its validity remains to be tested for hexagonal crystals. A strong (hkl) dependence of t (that is, non-isostress) in the polycrystalline specimen may partially or fully account for the observed lattice strain anisotropy. Consequently the elasticity tensor can no longer be uniquely determined, but is only partially constrained by the lattice strain anisotropy in equation (6) with trade-offs among the values of the C_{ij} and textural parameters. For example, development of the basal-plane slip texture common in h.c.p. metals²⁸ could conceivably lower the t of grains with their c axis at 45° orientation. In such a case, the present observations at ultrahigh pressures, combined with elastic anisotropy information from theoretical estimates or single-crystal ultrasonic measurements at lower pressures, would yield rheological information on single-crystal strength anisotropy and polycrystalline flow textures. Such information would affect our estimates of both elastic and anelastic anisotropies in the core²⁹.

Much information on bulk properties at high temperatures and pressures, and single-crystal elasticity and strength anisotropy, may be obtained by integrating techniques complementary to the three-dimensional RXD measurements reported here. These include ultrasonic studies, which provide accurate and direct determination of velocities below 20 GPa, hydrostatic X-ray diffraction, which provides lattice parameters and bulk moduli to multimegabar pressures, shock-wave experiments, which determine bulk elasticity along the high pressure-temperature Hugoniot, and *ab initio* calculations, which provide independent determinations of elasticity. The present integrated study reveals that the elasticity of the Earth’s inner core may represent the low shear modulus of h.c.p. Fe close to melting, or the existence of additional components with low shear-wave velocities but similar density to Fe. As well as perfectly aligned h.c.p. Fe, more complicated textures should be considered, including partial alignment of these phases. □

Received 24 March; accepted 17 September 1998.

1. Song, X. & Richards, P. G. Seismological evidence for differential rotation of the Earth’s inner core. *Nature* **382**, 221–224 (1996).
2. Su, W., Dziewonski, A. M. & Jeanloz, R. Planet within a planet: rotation of the inner core of Earth.

- Science* **274**, 1883–1887 (1996).
3. Glatzmaier, G. A. & Roberts, P. H. Rotation and magnetism of Earth’s inner core. *Science* **274**, 1887–1891 (1996).
4. Kuang, W. & Bloxham, J. An Earth-like numerical dynamo model. *Nature* **389**, 371–374 (1997).
5. Jephcoat, A. & Olson, P. Is the inner core of the Earth pure iron? *Nature* **325**, 332–335 (1987).
6. Stixrude, L. & Cohen, R. E. High-pressure elasticity of iron and anisotropy of Earth’s inner core. *Science* **267**, 1972–1975 (1995).
7. Dziewonski, A. M. & Gilbert, F. Solidity of the inner core of the Earth inferred from normal mode observations. *Nature* **234**, 465–466 (1971).
8. Masters, G. & Gilbert, F. Structure of the inner core inferred from observations of its spheroidal shear modes. *Geophys. Res. Lett.* **8**, 569–571 (1981).
9. Shearer, P. M., Toy, K. M. & Orcutt, J. A. Axi-symmetric Earth models and inner-core anisotropy. *Nature* **333**, 228–232 (1988).
10. Creager, K. C. Anisotropy of the inner core from differential travel times of the phases PKP and PKIKP. *Nature* **356**, 309–314 (1992).
11. Tromp, J. Support for anisotropy of the Earth’s inner core from free oscillations. *Nature* **366**, 678–681 (1993).
12. Mao, H. K., Wu, Y., Chen, L. C., Shu, J. F. & Jephcoat, A. P. Static compression of iron to 300 GPa and Fe₉₅Ni₅ alloy to 260 GPa: Implications for composition of the core. *J. Geophys. Res.* **95**, 21737–21742 (1990).
13. Brown, J. M. & McQueen, R. G. Phase-transitions, Grüneisen-parameter, and elasticity for shocked iron between 77 GPa and 400 GPa. *J. Geophys. Res.* **91**, 7485–7494 (1986).
14. Duffy, T. S. & Ahrens, T. J. in *High Pressure Research: Application to Earth and Planetary Sciences* (eds Syono, Y. & Manghnani, M. H.) 353–361 (Terra Scientific Publishing, Tokyo, 1992).
15. Singh, A. K., Mao, H. K., Shu, J. & Hemley, R. J. Estimation of single-crystal elastic moduli from polycrystalline x-ray diffraction at high pressure: Applications to FeO and iron. *Phys. Rev. Lett.* **80**, 2157–2160 (1998).
16. Li, B., Jackson, I., Gasparik, T. & Liebermann, R. C. Elastic wave velocity measurement in multi-anvil apparatus to 10 GPa using ultrasonic interferometry. *Phys. Earth Planet. Inter.* **98**, 79–91 (1996).
17. Hemley, R. J. *et al.* X-ray imaging of stress and strain of diamond, iron, and tungsten at megabar pressures. *Science* **276**, 1242–1245 (1997).
18. Singh, A. K. & Balasingh, C. The lattice strains in a specimen (hexagonal system) compressed nonhydrostatically in an opposed anvil high pressure setup. *J. Appl. Phys.* **75**, 4956–4962 (1994).
19. Ueda, T., Fujimori, N. & Yagi, T. Lattice strains in crystals under uniaxial stress field. *J. Appl. Phys.* **80**, 739–746 (1996).
20. Anderson, O. L., Isaak, D. G. & Yamamoto, S. Anharmonicity and the equation of state for gold. *J. Appl. Phys.* **65**, 1534–1543 (1989).
21. Jephcoat, A. P., Mao, H. K. & Bell, P. M. The static compression of iron to 78 GPa with rare gas solids as pressure-transmitting media. *J. Geophys. Res.* **91**, 4677–4684 (1986).
22. Simmons, G. & Wang, H. *Single Crystal Elastic Constants and Calculated Aggregate Properties* (MIT Press, Cambridge, MA, 1971).
23. Guinan, M. W. & Steinberg, D. J. Pressure and temperature derivatives of the isotropic polycrystalline shear modulus for 65 elements. *J. Phys. Chem. Solids* **35**, 1501–1512 (1974).
24. Söderlind, P., Moriarty, J. A. & Wills, J. M. First-principles theory of iron up to earth-core pressures: Structural, vibrational, and elastic properties. *Phys. Rev. B* **53**, 14063–14072 (1996).
25. Dziewonski, A. & Anderson, D. L. Preliminary reference earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
26. Tallon, J. The volume dependence of elastic moduli and the Born–Durand melting hypothesis. *Phil. Mag.* **39**, 151–161 (1979).
27. Singh, A. K., Balasingh, C., Mao, H. K., Hemley, R. J. & Shu, J. Analysis of lattice strains measured under non-hydrostatic pressure. *J. Appl. Phys.* **83**, 7567–7575 (1998).
28. Wenk, H. R., Takeshita, T. & Jeanloz, R. Development of texture and elastic anisotropy during deformation of hcp metals. *Geophys. Res. Lett.* **15**, 76–79 (1988).
29. Romanowicz, B., Li, X. D. & Durek, J. Anisotropy in the inner core: could it be due to low-order convection? *Science* **274**, 963–966 (1996).
30. Stacey, F. D. Theory of thermal and elastic properties of the lower mantle and core. *Phys. Earth Planet. Inter.* **89**, 219–245 (1995).

Acknowledgements. We thank J. Hu for technical help, L. Stixrude and R. E. Cohen for sharing theoretical data and discussions, T. Duffy for comments, and NSLS and APS for synchrotron beam time; the synchrotron facilities are supported by the DOE. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to H.M. (e-mail: MAO@gl.ciw.edu).

Human longevity at the cost of reproductive success

Rudi G. J. Westendorp*† & Thomas B. L. Kirkwood†

* Section of Gerontology and Geriatrics, Department of General Internal Medicine, and Clinical Epidemiology, Leiden University Medical Centre CO-P, PO Box 9600, 2300 RC Leiden, The Netherlands

† Biological Gerontology Group, Department of Geriatric Medicine and The School of Biological Sciences, University of Manchester, 3.239 Stopford Building, Oxford Road, Manchester M13 9PT, UK

The disposable soma theory on the evolution of ageing states that longevity requires investments in somatic maintenance that reduce the resources available for reproduction^{1,2}. Experiments in *Drosophila melanogaster* indicate that trade-offs of this kind exist in non-human species^{3–7}. We have determined the interrelationship between longevity and reproductive success in *Homo*

Table 1 Number of progeny and age at first childbirth dependent on the age at death of married aristocratic women

Age at death (yr)	Number	Proportion childless	Mean	Number of progeny (95% CI)	Mean	Age at first childbirth (yr) (95% CI)
<20	42	0.66	0.45	(0.29–0.69)	19.1	(15.9–22.5)
21–30	176	0.39	1.35	(0.86–2.11)	20.5	(19.4–21.6)
31–40	218	0.26	2.05	(1.33–3.18)	23.2	(22.3–24.2)
41–50	210	0.31	2.01	(1.30–3.11)	23.9	(22.9–24.9)
51–60	299	0.28	2.40	(1.56–3.71)	24.6	(23.8–25.4)
61–70	337	0.33	2.36	(1.53–3.63)	23.8	(23.0–24.6)
71–80	322	0.31	2.64	(1.71–4.07)	24.6	(23.8–25.3)
81–90	247	0.45	2.08	(1.35–3.24)	25.1	(24.1–26.1)
>90	57	0.49	1.80	(1.12–2.90)	27.0	(24.8–29.2)

Point estimates and 95% confidence intervals (95% CI) are adjusted for the trends over calendar time using Poisson regression (number of progeny) and linear regression (age at first childbirth), respectively.

sapiens using a historical data set from the British aristocracy. The number of progeny was small when women died at an early age, increased with the age of death, reaching a plateau through the sixth, seventh and eighth decades of life, but decreased again in women who died at an age of 80 years or over. Age at first childbirth was lowest in women who died early and highest for women who died at the oldest ages. When account was taken only of women who had reached menopause, who were aged 60 years and over, female longevity was negatively correlated with number of progeny and positively correlated with age at first childbirth. The findings show that human life histories involve a trade-off between longevity and reproduction.

Identifying the heritable contribution to variation in lifespan is an important component of understanding the biological basis of human ageing. An analysis of human life-history data should take into account the socioeconomic conditions affecting both longevity and reproduction. Instead of adjusting for socioeconomic characteristics, we primarily relied on restricting the population to members of British aristocratic families, which are reasonably homogeneous and in which social deprivation has not interfered unduly with the chance of a long life. Extensive genealogical records have been kept for aristocratic families dating back to as early as the year 740; these records have been published and computerized.

The associations between age at death and reproductive histories of women from the British aristocratic families are summarized in Table 1. The analysis of the families is further restricted to lifelong data from birth cohorts up to 1876, and to married women only.

For women who died young, the proportion who were childless was high (two-thirds), but nearly a third of all women and half of those who died above 80 years of age were also childless. As lifespan, number of progeny, and age of childbirth have drifted over calendar time, age-at-death-specific estimates of progeny number and age at first childbirth were adjusted for time trends by using multiple regression. The number of progeny was small if women died young, increased with age at death, reaching a plateau through the sixth, seventh and eighth decades of life, but declined again in women who lived 80 years or more. Mean age at first childbirth was lowest in women who died young, reached a plateau for those dying in middle age, and increased again for women who died at the oldest ages.

The increase in the number of progeny with age-at-death in women who lived up to middle age may reflect a longer time to bear offspring. Another likely possibility is that a small number of progeny and an early age of first childbirth among women who died early is the direct consequence of pregnancy-related mortality. We therefore performed a separate analysis by including only those women who had reached menopause. Among the women who lived to 60 years and over, the number of progeny was negatively correlated with female longevity (likelihood ratio, 24.2; d.f. = 3; $P < 0.001$). The age at first childbirth was positively correlated with longevity ($F = 4.08$; d.f. = 3, 516; $P = 0.007$). Similar statistically significant correlations were obtained when the cut-off was reduced from 60 to 50 years. To exclude the possibility that the association between longevity and late age of first childbirth was mediated simply through having fewer children, we also repeated the analysis adjusting for number of progeny and found that the results were unaffected (data not shown).

It is critical to ascertain that the associations between age at death and reproductive histories hold after potentially distorting demographic trends are taken into account. As expected, mortality rates for aristocratic women were highly dependent on age and calendar time (Fig. 1). The age-group-specific mortality rates show high child mortality, a steep decrease to a minimum at young adulthood, and a linear increase of mortality rate over middle age when plotted on a semilog scale. The slope over middle age, designated as the Gompertz coefficient, is an indication of the ageing process⁸. The decrease in mortality over calendar time occurred especially at young adulthood and middle age. Child mortality in the earliest birth cohorts may have been underestimated as not all newborns that died in remote centuries have come to the attention of genealogists. Despite the profound decrease of mortality, the slope of mortality rate at middle age was roughly similar. At the oldest ages, mortality rates of the early and late birth cohorts converged. This may reflect heterogeneity of the population, which is one of the explanations for deceleration of human mortality at older age, or simply the fact that old-age mortality did not start to decrease until the 1950s (ref. 9).

The decrease in mortality over calendar time underlies the increase in lifespan of the aristocratic women that began at

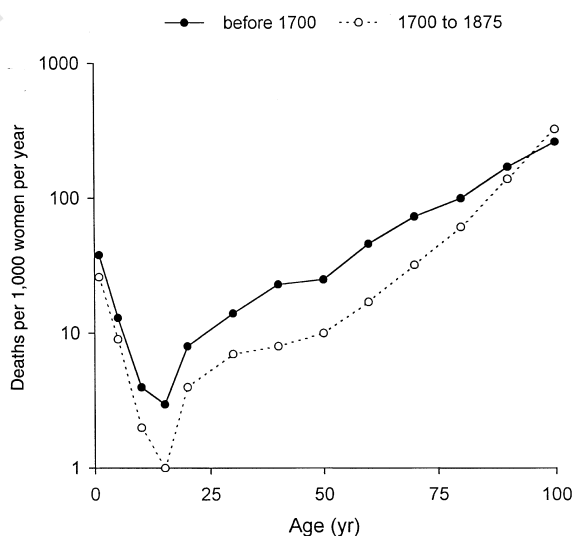


Figure 1 Mortality rates as a function of age for aristocratic women from different birth cohorts, plotted on a semilog scale.

Table 2 Age at death and number of progeny in aristocratic women over calendar time

Birth cohort (yr)	Number	Proportion married	Age at death (yr)		Number of progeny*	
			Mean	(95% CI)	Mean	(95% CI)
<1500	337	0.89	46.1	(44.1–48.3)	2.34	(2.18–2.52)
1501–1550	106	0.83	42.9	(38.6–47.2)	2.31	(1.98–2.70)
1551–1600	122	0.74	44.7	(40.7–48.8)	2.03	(1.73–2.39)
1601–1650	190	0.79	45.8	(42.2–49.4)	2.83	(2.51–3.19)
1651–1700	296	0.68	43.1	(40.1–46.1)	2.41	(2.15–2.71)
1701–1725	124	0.61	43.3	(38.5–48.2)	2.22	(1.88–2.63)
1726–1750	120	0.73	52.3	(47.7–56.9)	2.50	(2.15–2.91)
1751–1775	149	0.82	53.1	(48.8–57.4)	2.19	(1.90–2.52)
1776–1800	174	0.82	55.3	(51.3–59.3)	2.20	(1.92–2.51)
1801–1825	222	0.83	60.5	(57.5–63.6)	1.76	(1.54–2.01)
1826–1850	216	0.83	61.4	(58.0–64.7)	2.06	(1.82–2.34)
1851–1875	385	0.74	68.0	(65.7–70.4)	1.54	(1.37–1.73)

* Calculated for married women only.

around 1700 (Table 2). The mean age at death for women remained at about 45 until the first decades of the eighteenth century. Thereafter, lifespan steadily increased to a mean of 68 years for women born between 1850 and 1875. The average number of progeny for a married woman was 2.3 before 1500 and at a maximum of 2.8 at the beginning of the seventeenth century (during the reign of Queen Elizabeth I). The average number of progeny then steadily decreased to a minimum of 1.5 for the 1850–1875 birth cohort. This demographic transition among the British aristocracy (that is, the increase in lifespan and concomitant decrease in number of progeny) began some 150 years earlier than in the general population¹⁰.

As a consequence of the demographic transition, it may also be expected that the relation between age at death and reproductive histories altered at around this time. This effect is revealed in Fig. 2, which shows the average progeny number plotted against the mother's age at death for the periods before and after 1700. In married women born before 1700, the average number of progeny was generally high and the decrease in progeny number among those who died at the oldest ages was pronounced. None of the six women who were born before 1700 and who reached the exceptional age of 90 years and over had more than two children. Among women born after 1700, the average number of children was generally lower and the decrease in the number of progeny among those who died at the oldest ages was less obvious. However, when only postmenopausal women (aged 60 years and over) were included in the analysis, the decrease in the number of progeny with the age at death was statistically significant for both the pre-

1700 period (likelihood ratio, 21.4; d.f. = 3; $P < 0.001$) and the post-1700 period (likelihood ratio, 12.9; d.f. = 3; $P = 0.005$). The downward trend in the mean number of progeny with high age at death remained apparent even when all childless women were excluded from the analysis (data not shown). The mother's age of first childbirth increased significantly with age at death from 60 years onwards in the pre-1700 period ($F = 3.01$; d.f. = 3, 161; $P = 0.032$) and the post-1700 period ($F = 4.30$; d.f. = 3, 352; $P = 0.005$). The results are unlikely to be distorted by underreporting of child mortality as incompleteness of the genealogies is not likely to be different for women who died in their seventh or eighth decade from that for women who lived to 80 years and over.

We also studied the association between the number of progeny and lifespan in males. Analysis of the data for the males revealed essentially the same pattern of association between fecundity and longevity as for the females (Table 3). We found evidence for both sexes that lifespan itself is heritable. Empirical heritability^{11,12} estimates for longevity of daughters dependent on the age at death of the mother and the father were 0.37 ± 0.07 (mean $\pm$ standard error) and 0.22 ± 0.07 , respectively, whereas for sons they were 0.16 ± 0.06 and 0.20 ± 0.04 . All heritability coefficients were statistically significant ($P < 0.01$). The estimates of paternal heritability are nearly identical to previously published estimates for families of comparable socioeconomic status¹³.

The relation between age at death and progeny number in British aristocratic women (and men) is consistent with the hypothesis that the longest lived individuals have reduced fertility compared with the majority of the population. This effect is seen most strongly in women born before 1700, for whom the number of children was larger than for women born between 1700 and 1875. The demographic transition towards increased lifespan probably accounts for the reduction in progeny number after this time. If children came to have a greater expectation of life, the pressure on the number of progeny would have been reduced, as has occurred subsequently through much of the world as a consequence of economic development. Once the number of children began to fall, the difference in

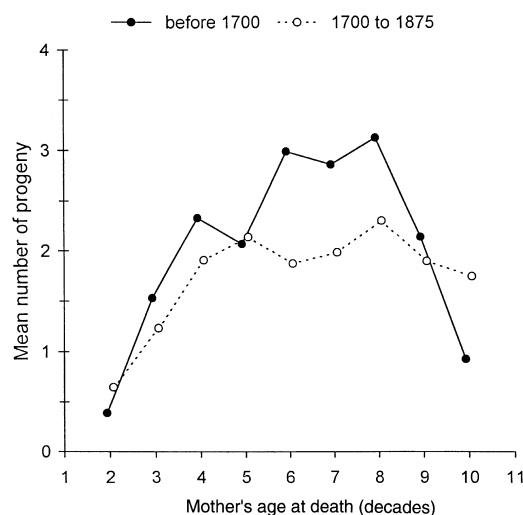


Figure 2 Point estimates of progeny number for married aristocratic women from different birth cohorts as a function of age at death. The estimates of progeny number are adjusted for trends over calendar time using multiple regression.

Table 3 Number of progeny dependent on the age at death of married aristocratic men

Age at death (yr)	Birth cohorts before 1700		Birth cohorts 1700 to 1875	
	No. of males	Mean (95% CI) progeny	No. of males	Mean (95% CI) progeny
< 20	23	0.43 (0.23–0.81)	3	0.34 (0.20–0.56)
21–30	121	1.19 (0.62–2.25)	50	1.00 (0.49–2.05)
31–40	198	1.90 (1.02–3.57)	163	1.46 (0.91–2.36)
41–50	308	2.38 (1.28–4.45)	328	1.82 (1.14–2.91)
51–60	366	2.96 (1.59–5.51)	476	2.41 (1.51–3.84)
61–70	341	3.10 (1.66–5.78)	801	2.19 (1.37–3.48)
71–80	206	2.97 (1.59–5.55)	987	2.26 (1.42–3.60)
81–90	68	2.40 (1.26–4.54)	571	2.29 (1.44–3.64)
> 90	13	1.42 (0.65–3.08)	96	2.06 (1.27–3.33)

Point estimates and 95% confidence intervals are adjusted for the trends over calendar time using Poisson regression.

fertility between long-lived individuals and the general population became less pronounced but remained statistically significant.

In spite of our restriction of the population to British aristocratic families in order to limit socioeconomic diversity, environmental factors as well as genetic factors could explain the correlation between longevity and fertility. For example, a favourable environment for longevity may be one in which reproductive pressures are also reduced. We tested this idea using the spouse's age at death as a proxy of the marital environment relevant for longevity. If environment was favourable for longevity, similar effects should be seen for both marriage partners. However, when the husband's age at death was entered as an explanatory variable into the regression models, neither the association for women between progeny number and age at death nor its statistical significance was altered (data not shown). This strengthens the interpretation that the decrease in progeny number in long-lived women has its basis in evolutionary genetics.

The evolutionary theories suggest that the origins of human ageing lie in the observations that the force of natural selection progressively weakens with increasing adult age^{14–16}, and that the acquisition of longevity involves significant metabolic costs in the form of investments in somatic maintenance or 'longevity assurance' mechanisms^{1,2}. The first observation suggests that genes with late deleterious effects can accumulate within the gene pool. The second suggests that there may be a trade-off between reproductive success and longevity, because resources invested in longevity assurance may be at the expense of reproduction. These mechanisms are not mutually exclusive, and generalized trade-offs between early and late life fitness are also implied by the pleiotropic genes hypothesis¹⁵.

In the fruitfly *Drosophila melanogaster*, there is evidence for the general existence of trade-offs between longevity and reproduction for females and males^{3–7}. A selection regime favouring flies that retained fertility at later ages resulted in populations with increased lifespans, reduced fertility early in life³, and enhanced resistance to a variety of stresses¹⁷, suggesting that the mechanisms underlying the increase in lifespan involved greater investments in somatic durability. Direct selection for longevity exploiting the dependence on temperature of the fruitfly lifespan also produced long-lived populations with significantly reduced fertility⁴, underpinning a genetic interpretation of the trade-off. However, non-genetic correlations between longevity and fecundity in *Drosophila* have also been described^{5–7}. In particular, both male and female flies prevented from mating show increased lifespans.

The data for humans presented here, and the increased likelihood that a woman becomes a centenarian when she bears children at a later age (in her forties)¹⁸, are reminiscent of the *Drosophila* experiments and support the predictions of evolutionary theories of ageing. In spite of the caveats in analysing and interpreting human life-history data, records for British aristocrats between the eighth and nineteenth centuries are compatible with the hypothesis that the human life history has a heritable component that involves a trade-off between fertility and longevity. This is masked in contemporary populations by a reduced birth rate, but may nevertheless be important for genetic determinants of longevity. □

Methods

As our data source we used the tenth edition of the Peerage CD (S&N Genealogy Supplies, Salisbury). This CD-ROM provides demographic information on 33,497 individuals (19,380 males and 13,667 females) and 18,125 marriages. The database was compiled by John Bloore, based on a test file provided with the genealogical program PEDIGREE that contained about 250 individuals, ancestors of the present Queen of England. As the program grew in size, a database was first created for the ancestors of the Prince of Wales (5,499 individuals) and was then expanded to include the lineages of the hereditary British peers. The database, alphabetically registered, now spans the Dukes and Earls of Abercorn to the Barons of Willoughby de Broke. Because of

extensive connections between the British peerage and aristocratic families in the rest of Europe, about 15% of the entries refer to individuals outside the UK, mostly in Germany and France.

The genealogical data are based primarily on refs 19–25. The first marriage recorded took place in 740, the last in 1995. Details are provided on the dates of birth or baptism, dates of marriage and death, the identity of parents and spouse, and the number of children. Some dates of birth and death were missing. Data were missing more frequently for women than for men, probably as a side-effect of the predominantly patrilineal inheritance pattern. To study only birth cohorts for which longevity records would be as complete as possible, we restricted the analysis to individuals born before 1876. The raw data were imported as a text file and checked for obvious errors in the database such as negative age or children born after the death of the mother. Individuals with obviously erroneous records were deleted from the data set; the number of deleted records made up less than 1% of the total.

The mean number of progeny per woman, defined as the fertility rate, and the mean age at first childbirth were calculated for ever-married women within 10-year age groups based on the women's age at death. The database did not contain any records of children born to unmarried women. Point estimates and 95% confidence intervals were obtained by cross-tabulation. Because mean age at death and the average number of progeny changed over calendar time, estimates were adjusted for trends over calendar time by using Poisson regression (number of progeny) and linear regression (mean age at first childbirth). Calendar time, based upon an individual's date of birth, was entered into the models as a single category for births before 1500, four 50-year categories for births between 1500 and 1700, and seven 25-year categories for births between 1700 and 1875, where the records were most dense. Statistical analysis was performed using the EGRET and SPSS software. Significant differences between categories were determined using the likelihood-ratio statistic (Poisson models) and the *F*-test (linear models) with the value of α set at 0.05. Estimates of the coefficients of empirical heritability were determined as twice the regression coefficients from a linear regression of offspring from age 30 years onwards on parents¹¹, after adjustment for trends over calendar time.

Received 2 October; accepted 28 October 1998.

1. Kirkwood, T. B. L. Evolution of ageing. *Nature* **270**, 301–304 (1977).
2. Kirkwood, T. B. L. & Rose, M. R. Evolution of senescence: late survival sacrificed for reproduction. *Phil. Trans. R. Soc. Lond. B* **332**, 15–24 (1991).
3. Partridge, L. & Barton, N. H. Optimality, mutation and the evolution of ageing. *Nature* **362**, 305–311 (1993).
4. Zwaan, B. J., Bijlsma, R. & Hoekstra, R. F. Direct selection of life span in *Drosophila melanogaster*. *Evolution* **49**, 649–659 (1995).
5. Fowler, K. & Partridge, L. A cost of mating in female fruitflies. *Nature* **338**, 760–761 (1989).
6. Prowse, N. & Partridge, L. The effects of reproduction on longevity and fertility in male *Drosophila melanogaster*. *J. Insect Physiol.* **43**, 501–512 (1997).
7. Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. Cost of mating in *Drosophila melanogaster* is mediated by male accessory-gland products. *Nature* **373**, 241–244 (1995).
8. Finch, C. E. *Longevity, Senescence, and the Genome* (Univ. Chicago Press, Chicago, 1990).
9. Vaupel, J. W. et al. Biodemographic trajectories of longevity. *Science* **280**, 855–860 (1998).
10. Hollingwood, T. H. Mortality in the British peerage families since 1600. *Population* **18** (suppl.), 323–351 (1977).
11. Khoury, M. J., Beaty, T. H. & Cohen, B. H. *Fundamentals of Genetic Epidemiology* (Oxford Univ. Press, New York, 1993).
12. Finch, C. E. & Tanzi, R. E. Genetics of ageing. *Science* **278**, 407–411 (1997).
13. Gavrilova, N. S. et al. Evolution, mutations, and human longevity: European royal and noble families. *Hum. Biol.* **70**, 799–804 (1998).
14. Medawar, P. B. *An Unsolved Problem of Biology* (Lewis, London, 1952).
15. Williams, G. C. Pleiotropy, natural selection and the evolution of senescence. *Evolution* **11**, 398–411 (1957).
16. Charlesworth, B. *Evolution in Age-structured Populations* (Cambridge Univ. Press, Cambridge, 1994).
17. Service, P. M., Hutchinson, E. W., MacKinlay, M. D. & Rose, M. R. Resistance to environmental stress in *Drosophila melanogaster* selected for postponed senescence. *Physiol. Zool.* **58**, 380–389 (1985).
18. Perls, T. T., Alpert, L. & Fretts, R. C. Middle-aged mothers live longer. *Nature* **389**, 133 (1997).
19. Townsend, P. *Burke's Genealogical and Heraldic History of the Peerage Baronetage and Knightage* (Burke's Peerage, London, 1963).
20. Gibbs, V., Doubleday, H. A., De Walden, H. & White, G. H. *The Complete Peerage of England, Scotland, Ireland, Great Britain and the United Kingdom* (St Catherine, London, 1910–1959).
21. Kidd, C. & Williamson, D. *Debrett's Peerage and Baronetage* (Macmillan, London, 1994).
22. Paget, G. *The Lineage and Ancestry of HRH Prince Charles, Prince of Wales* (Skilton, Edinburgh, 1977).
23. *Whittaker's Almanack* (Whittaker, London, 1996).
24. Moncreiffe, I. *Royal Highness, Ancestry of the Royal Child* (Hamish Hamilton, London, 1982).
25. Morby, J. E. *The Wordsworth Handbook of Kings and Queens* (Wordsworth Editors, Garsme, 1994).

Acknowledgements. This work was supported by NESTOR. We thank L. Gavrilova and N. Gavrilova for identifying the database, J. Bloore for computerizing the genealogical data, and J. Vandenbroucke for critical comments on the manuscript.

Correspondence and requests for materials should be addressed to R.G.J.W. (e-mail: RudiWestendorp@compuserve.com).

Role of spectral detail in sound-source localization

Abhijit Kulkarni & H. Steven Colburn

Hearing Research Center and Department of Biomedical Engineering,
Boston University, Boston, Massachusetts 02215, USA

Sounds heard over headphones are typically perceived inside the head¹ (internalized), unlike real sound sources which are perceived outside the head (externalized). If the acoustical waveforms from a real sound source are reproduced precisely using headphones, auditory images are appropriately externalized and localized¹⁻⁴. The filtering (relative boosting, attenuation and delaying of component frequencies) of a sound by the head and outer ear provides information about the location of a sound source by means of the differences in the frequency spectra between the ears as well as the overall spectral shape. This location-dependent filtering is explicitly described by the head-related transfer function (HRTF) from sound source to ear canal. Here we present sounds to subjects through open-canal tube-phones and investigate how accurately the HRTFs must be reproduced to achieve true three-dimensional perception of auditory signals in anechoic space. Listeners attempted to discriminate between 'real' sounds presented from a loudspeaker and 'virtual' sounds presented over tube-phones. Our results show that the HRTFs can be smoothed significantly in frequency without affecting the perceived location of a sound. Listeners cannot distinguish real from virtual sources until the HRTF has lost most of its detailed variation in frequency, at which time the perceived elevation of the image is the reported cue.

Although previous studies¹⁻⁴ showed that real and virtual sounds (matched to have identical acoustic waveforms in the ear canals) were indistinguishable when the subject was wearing headphones, they did not address the question of the sensitivity of the subjects to the details of the HRTF. Here we presented virtual sounds to subjects' ears through open-canal 'tube-phones' (Fig. 1), so that 'real' sounds were essentially unaffected by the presence of the tube-phones and sounds from real and virtual sources could be directly compared. The paired-comparison experiments required listeners simply to report the order of the stimuli: real first or virtual first. Subjects were given practice and trial-by-trial feedback so that small

differences in the location or the realism of the sounds could be used for judgements. In the measurements with smoothed HRTFs, the stimulus spectrum was randomized so that non-spatial attributes (such as timbre) could not be used for discrimination.

In the first experiment, four subjects were tested for their ability to distinguish natural stimuli from virtual stimuli, which were constructed to match the natural stimuli exactly. Source spectra were not varied (beyond the stochastic nature of the noise waveforms) in these initial validation experiments. We tested four azimuthal locations (0, 45, 135 and 180 degrees, where 0 degrees is straight ahead) in separate sets of trials. None of the subjects was able to distinguish natural from virtual stimuli. Performance of each subject at each location was within the bounds expected from chance performance. (The percentages of correct values obtained are plotted with the data from the smoothing experiment in Fig. 2.) These results were consistent with the subjects' impressions that they perceived both types of stimuli as completely natural and were unable to perceive any differences between the virtual and the free-field stimuli. These basic results show the adequacy of our methods for measurements of HRTFs and for virtual-stimulus presentation as well as the naturalness of the resulting perceptions.

In a second set of experiments, we found that, provided that the overall interaural time delay (ITD) was maintained to be consistent with that in the natural sound waveforms, virtual sounds synthesized through HRTFs with the original magnitude spectra (that is, that of the natural stimulus) but simplified phase spectra were also indistinguishable from the natural sounds. This lack of dependence of the spatial percept on the frequency-dependent detail of the ITD is consistent with previous results^{1,5,6} and led us to develop our procedure in which the modified HRTFs were given the minimum-phase response⁷ for each magnitude, supplemented by a pure ITD that was calculated to make the ITD of the virtual stimulus match the ITD of the real stimulus.

In all of our other experiments, the magnitude spectra of empirical HRTFs measured from individual listeners were systematically smoothed and the discriminability of the resulting virtual stimuli from free-field stimuli was assessed. This smoothing was performed by expressing the HRTF log-magnitude spectrum as a Fourier series and reconstructing the HRTF spectrum using a truncated series (see Methods). The nature of the smoothing operation is shown in Fig. 3 for a representative HRTF for six values of the smoothing parameter. For these experiments, the signal components of the test stimuli (the source waveforms) were

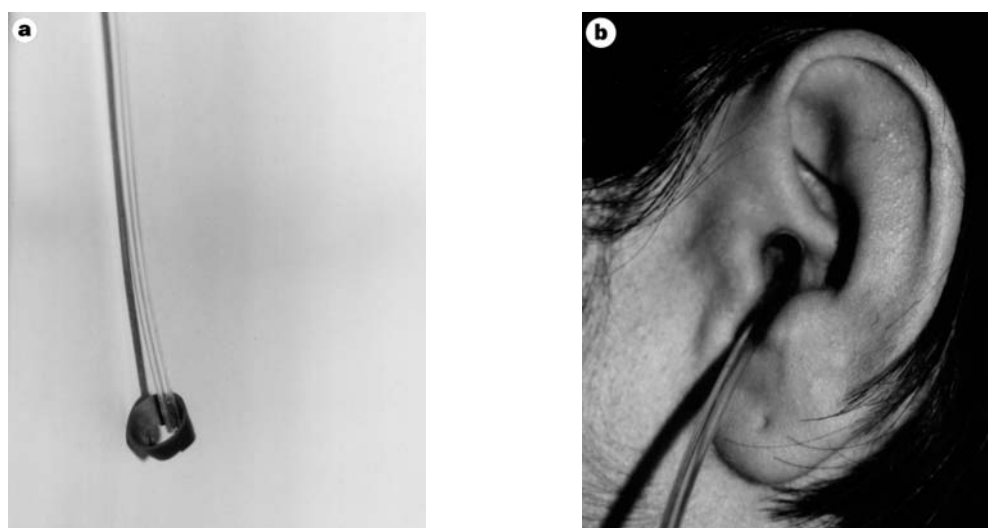


Figure 1 The tube-phone apparatus. **a**, The apparatus attached to a customized latex-rubber hollow shell to fit the ear canals of listeners. **b**, The apparatus

inserted in the ear canal of a listener. Note that the shell provides a snug fit flush with the walls of the ear canal with minimal distortion of ear-canal acoustics.

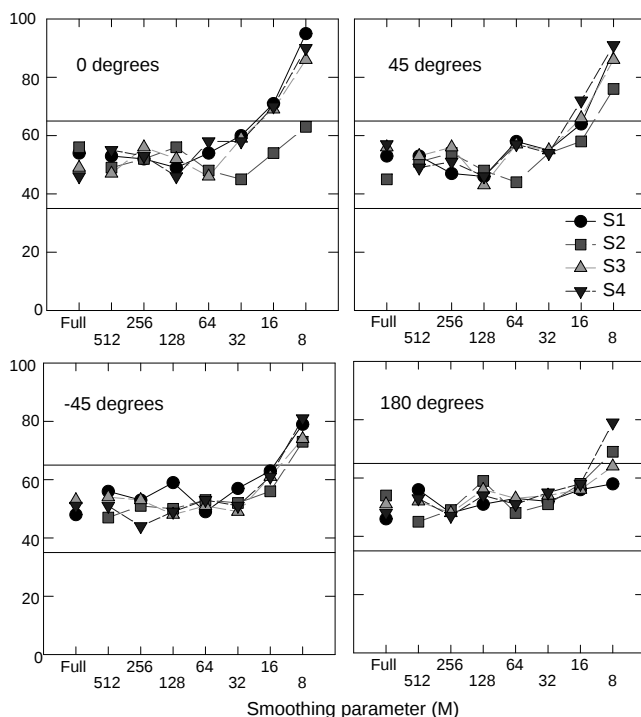


Figure 2 Discrimination performance of four subjects (S1–S4) as a function of the smoothing parameter, M (see Methods), at each azimuthal position tested. The 95% confidence bounds for chance performance are shown by the two horizontal lines in each graph. The left-most points in each panel (marked 'Full' along the abscissa) correspond to performance in response to the complete HRTF and no stimulus randomization.

randomly varied in each trial. The interval-to-interval variation (see Methods) in the stimulus waveform introduced an uncertainty in the quality of the sounds in each presentation and prevented listeners from using non-spatial differences in sound quality (such as timbre) to discriminate between natural and virtual auditory images. We determined a priori that this variation did not alter the location of the image from a loudspeaker.

The performance of the four subjects tested is shown in Fig. 2 for four azimuthal locations (0, 45, 135 and 180 degrees). The performance (in per cent correct) is plotted as a function of the smoothing parameter (the number of Fourier coefficients used in the HRTF reconstruction). Performance was largely unaffected by the spectral smoothing; performance was always within the bounds of chance performance for reconstructions having more than 16 coefficients. In fact, performance was seldom outside these bounds except at the extreme smoothing condition (8 Fourier coefficients). Comparisons of Figs 2 and 3 show that discrimination performance remained within chance bounds for fairly large amounts of smoothing and became consistent only at the extreme smoothing conditions tested.

The subjective reports of all listeners were consistent with the results reported above. All subjects reported hearing both free-field and virtual stimuli from the free-field loudspeaker in almost all trials. It was more surprising, however, that, even at the extreme smoothing condition (8 Fourier coefficients) when the discrimination performance of subjects was high, all subjects reported complete externalization of the virtual sound image. In fact, all subjects reported that the only useful cue in the extreme smoothing conditions was the elevation of the virtual image (to varying degrees) above the free-field sound image. The elevation of the virtual image in the extreme smoothing conditions is consistent with the observation that the magnitude spectra of HRTFs from high elevations are relatively smooth compared with those from lower elevations. The

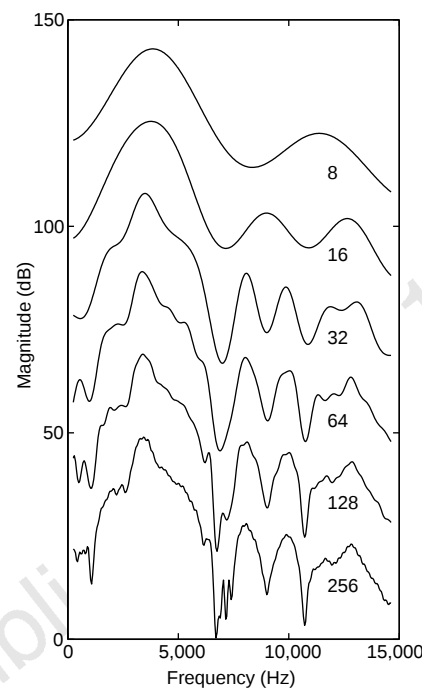


Figure 3 Spectral smoothing of the left-ear HRTF magnitude spectra measured from a representative subject (S1) for the 0-degree location. The HRTFs are shown on a relative scale and are offset from each other by 20 dB. The smoothing parameter, M , is described in the Methods and is indicated next to each curve.

compelling perception of a well-externalized sound was also perceived by several experienced psycho-acousticians using spectrally smooth stimuli presented through the tube-phone apparatus. All of our observations are consistent with the conclusion that spectral detail is not important for sound externalization and localization.

Our results show that, at least in anechoic space when no head movements occur and when the visual cues are consistent and accurate, details of the HRTFs (magnitude and phase) are not responsible for a well-localized, externalized perception of a sound source. These conclusions do not indicate that individual variations in the HRTF are unimportant, as we used individualized HRTFs in all of our experiments. Our results are consistent with studies in which crude approximations to the natural ear-input signals were perceived as natural provided that these waveforms were made to change in a manner consistent with the movement of the listener's head⁸. We suggest that a critical factor contributing to the success of our experimental technique is the playback apparatus: the tube-phone system preserves the acoustics of a natural listening condition. Specifically, both the pathways for internally and externally generated sounds and the ear-canal impedances are consistent with the natural (open-ear) listening condition. In contrast to suggestions of previous studies^{1,2,4} that quality deteriorates when HRTFs are not accurately reproduced, our results show that the fine structure of the HRTF is relatively unimportant for auditory spatial attributes, including externalization. □

Methods

Tube-phone apparatus. Virtual field stimuli were presented through the ER-2 tube-phone (Etymotic Research) which was used without the provided foam-tips. The use of the tube-phones without foam-tips results in a second-order (40 dB per decade), low-frequency (below 2 kHz) roll-off in the frequency response; we compensated for this roll-off by applying a second-order pre-emphasis filter over the relevant frequency region. The virtual acoustical signals

were computer-generated using individualized HRTFs measured from the entrance of the blocked ear canal (for reasons of repeatability and good signal-to-noise ratio⁹). A position-independent transformation was then applied to these blocked-canal HRTFs to make the pressure at the eardrum of the listener during playback consistent with the free-field pressure at that point. Natural-sound field stimuli were presented from a loudspeaker which was located 1.15 m from the listener's head. Both the real and the natural stimuli had a bandwidth of 350–15,000 Hz.

The presence of the 1.35-mm-diameter tubes in the ear canals had negligible effects on the natural sound field. This was observed subjectively, and confirmed by acoustical measurements and by a study comparing localization performance in the median-sagittal plane with and without the tube-phones present. We also measured the acoustical crosstalk between the tube-phones at the two ears to be negligible at the presentation levels used (50 dB sound pressure level (SPL) measured with steady-state excitation in the ear canal of the KEMAR acoustical mannequin).

Psychophysical discrimination paradigm. The experimental paradigm was a two-interval, two-alternative, forced-choice (2I, 2AFC) discrimination task. Each interval consisted of a pair of 80-ms noise bursts (having 10-ms raised-cosine rise/fall times) presented at 45 dB SPL and separated by 100 ms. The time between the two intervals of each trial was 200 ms. The stimuli in each interval were either synthesized virtually or delivered from a free-field loudspeaker in an anechoic chamber (from the corresponding location), in random order. In the experiments in which the HRTF spectrum was smoothed, the signal components of the noise stimuli were randomized in 1/3-octave bands by ± 5 dB in each trial to prevent listeners from using any non-spatial differences in sound quality in discriminating the virtual sounds from the real sounds. In a localization experiment conducted in the median-sagittal plane we determined that the localization performance of listeners was not significantly different with or without the spectral rove in the noise stimulus.

The location of a single source was tested during each run and feedback regarding the correct answer was provided in each trial. HRTFs were measured from listeners at the start of each experimental run with head position monitored by a Polhemus head-position sensor (which was worn on top of the subject's head, held by a plastic frame). A computer program instructed the subject to find 0-degree yaw, pitch and roll coordinates; the program read the Polhemus sensor output and delivered voiced instructions (consisting of digitized speech samples) over a loudspeaker. The instructions consisted of the words 'turn', 'roll', 'left', 'right', 'up', 'down', and 'hold it'. The measurement was only made when the reference position had been reached. The subject remained seated in the same location for the rest of the session and was prompted (if necessary) to achieve the same orientation for each presentation during playback. Head position was constantly monitored during the course of the stimulus playback and the trial was automatically terminated if any motion was detected. This ensured that discrimination between the stimuli did not result from any misalignment between the listener and the speaker and that no dynamical cues were used to determine the source of the stimulation.

HRTF smoothing. The discrete Fourier transform, $H(e^{j\omega})$, of an HRTF impulse response can be expressed as a discrete sequence, $H[k]$, where $H[k] = H(e^{j\omega})|_{\omega = (2\pi/N)k}$, where N is the number of sample values. The discrete realization of the log-magnitude spectrum may be expressed by the Fourier series:

$$\hat{H}[k] = \log|H[k]| = \sum_{n=0}^{N/2} C(n)\cos(2\pi nk/N)$$

where $C(n)$ is the n th coefficient in the Fourier series for the sequence $\log|H[k]|$.

By using a partial Fourier series of degree $M < N/2$, a smoothed fit to the HRTF spectrum can be obtained:

$$\hat{H}[k] = \sum_{n=0}^M C(n)\cos(2\pi nk/N)$$

The choice of $M = N/2$ corresponds to an exact reconstruction of the HRTF magnitude spectrum and the choice of $M = 1$ corresponds to a reconstruction having a flat spectrum at the average value of the HRTF.

For an empirical HRTF impulse response having N sample values, a Fourier series having $M = N/2$ terms provides an exact reconstruction of the empirical magnitude spectrum. The empirical HRTF in our experiments consisted of

1,024 time points (and thus correspond to a Fourier series with 512 terms). The smoothed magnitude spectra used in our experiments were constructed using 256, 128, 64, 32, 16 and 8 Fourier coefficients. The HRTFs were implemented as minimum-phase filters, augmented with a frequency-independent time delay, which was consistent with the overall ITD in the empirical HRTF measurements for the location being tested.

Received 10 August; accepted 23 October 1998.

1. Hartmann, W. M. & Wittenberg, A. On the externalization of sound images. *J. Acoust. Soc. Am.* **99**, 3678–3688 (1996).
2. Wenzel, E. M. Localization in virtual acoustic displays. *Presence Teleop. Virt. Environ.* **1**, 80–107 (1992).
3. Wightman, F. L. & Kistler, D. J. Headphone stimulation of free-field listening I: stimulus synthesis. *J. Acoust. Soc. Am.* **85**, 858–867 (1989).
4. Zahorik, P., Wightman, F. L. & Kistler, D. J. in *Proc. Acoustics, Speech and Signal Processing (Institute of Electrical and Electronics Engineers) Workshop Appl. Signal Processing Audio Acoustics* (IEEE Press, New York, 1995).
5. Kistler, D. J. & Wightman, F. L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.* **91**, 1637–1647 (1991).
6. Kulkarni, A., Isabelle, S. K. & Colburn, H. S. Sensitivity of human subjects to head-related transfer function phase spectra. *J. Acoust. Soc. Am.* (submitted).
7. Oppenheim, A. V. & Schaffer, R. W. *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
8. Loomis, J. M. & Soule, J. I. in *Proc. Society for Information Display 1996 Int. Symp.* 965–968 (Society for Information Display, San Jose, CA, 1996).
9. Shaw, E. in *Binaural and Spatial Hearing in Real and Virtual Environments* (eds Gilkey, R. H. & Anderson, T. R.) 25–48 (Lawrence Erlbaum, Mahwah, NJ, 1997).

Correspondence and requests for materials should be addressed to A.K. (e-mail: abhijit@enga.bu.edu).

Hedgehog stimulates maturation of *Cubitus interruptus* into a labile transcriptional activator

Johanna Talavera Ohlmeyer & Daniel Kalderon

Department of Biological Sciences, Columbia University,
1212 Amsterdam Avenue, New York, New York 10027, USA

In *Drosophila*, signalling by the protein Hedgehog (Hh) alters the activity of the transcription factor *Cubitus interruptus* (Ci) by inhibiting the proteolysis of full-length Ci (Ci-155) to its shortened Ci-75 form^{1,2}. Ci-75 is found largely in the nucleus and is thought to be a transcriptional repressor¹, whereas there is evidence^{3–5} to indicate that Ci-155 may be a transcriptional activator^{1,2,6}. However, Ci-155 is detected only in the cytoplasm, where it is associated with the protein kinase Fused (Fu), with Suppressor of Fused (Su(fu)), and with the microtubule-binding protein Costal-2 (refs 1,7–9). It is not clear how Ci-155 might become a nuclear activator. We show here that mutations in *Su(fu)* cause an increase in the expression of Hh-target genes in a dose-dependent manner while simultaneously reducing Ci-155 concentration by some mechanism other than proteolysis to Ci-75. Conversely, eliminating Fu kinase activity reduces Hh-target gene expression while increasing Ci-155 concentration. We propose that Fu kinase activity is required for Hh to stimulate the maturation of Ci-155 into a short-lived nuclear transcriptional activator and that *Su(fu)* opposes this maturation step through a stoichiometric interaction with Ci-155.

Hh secreted from posterior compartment cells of the wing disc in *Drosophila* induces expression of *decapentaplegic* (*dpp*) in anterior cells as far as the site of the future vein 3 (refs 10, 11) and anterior *engrailed* (*en*) expression in a narrower strip of cells close to the anterior–posterior (AP) compartment border^{11,12} (Fig. 1a). The amount of Dpp protein secreted from the AP border dictates the position at which vein 2 forms¹³. En and other factors induced by Hh contribute to the formation of vein 3 and associated campaniform sensillae at the anterior edge of the Hh-signalling territory¹⁴.

These effects of Hh signalling can be reproduced in more anterior regions by ectopic expression of Hh¹⁰, or by genetic inactivation of the Hh-receptor Patched (Ptc) in somatic clones² (Fig. 1a–c; Table 1). Anterior clones deficient in protein kinase A (PKA) activity also induce ectopic expression of Hh-target genes and reorganize wing pattern² (Fig. 1e; Table 1).

Elimination of *Su(fu)* activity moderately increased the effects of *ptc* and *PKA* mutant clones, enhancing the displacement of vein 2 from the anterior edge of the mutant clone and the number of ectopic sensillae induced (Table 1). A similar increase in sensillae and the distance between vein 2 and vein 3 was produced by *Su(fu)* mutations in otherwise wild-type wings, indicating a possible effect on normal Hh signalling (Table 1). Transcriptional responses were also affected. Ectopic *En* expression was induced less strongly in anterior *PKA* mutant clones than in *ptc* mutant clones (Fig. 2a, g), but was increased by loss of *Su(fu)* only in *PKA* mutant clones to reach the same level as in *ptc* clones (Fig. 2c, i).

A more marked enhancement of *PKA* mutant phenotypes by *Su(fu)* mutation was seen in embryos. In stage-9–12 embryos, *wingless* (*wg*) is normally expressed in one cell-wide stripes in each segment: ectopic Hh protein or *ptc* mutations induce an anterior expansion of the *wg* stripes^{5,15}, whereas inhibition of PKA causes

very little or no ectopic *wg* expression¹⁶. Loss of *Su(fu)* activity greatly enhanced induction of ectopic *wg* RNA by a dominant PKA inhibitor expressed in roughly alternating segments (Fig. 3a, c).

Hh signalling or loss of *ptc* or *PKA* activity are each accompanied by an increase in the amount of Ci-155 that can be detected *in situ* with an antibody that recognizes Ci-155 but not Ci-75 (refs 1, 2, 17) (Fig. 2b, h). Surprisingly, *Su(fu)* mutations decreased the amount of Ci-155 induced by Hh or by loss of *ptc* or *PKA* activity. Ci-155 staining at the AP border (Fig. 5b) was greatly reduced in *Su(fu)* wing discs (Fig. 5f). Ci-155 concentrations were only slightly increased in *PKA* mutant clones in *Su(fu)* discs (Fig. 2j) and no increase of Ci-155 was detected in *ptc* mutant clones in *Su(fu)* discs (Fig. 2d). Inhibition of PKA in embryos increased the amount of Ci-155 (ref. 16) (Fig. 3b), but not in the absence of *Su(fu)* activity (Fig. 3d).

Western blots of wild-type and *Su(fu)* mutant embryo and wing-disc extracts confirmed that *Su(fu)* mutations reduced the accumulation of Ci-155 (Fig. 4). Furthermore, PKA inhibition greatly increased the amount of Ci-155 at the expense of Ci-75 in embryos, whereas loss of both PKA and *Su(fu)* activity reduced Ci-155 to below the level in wild-type embryos. The reduction in Ci-155 caused by *Su(fu)* mutation was never accompanied by an increase in Ci-75, indicating that proteolysis of Ci-155 to Ci-75 was not responsible.

Conversion of Ci-155 to Ci-75 requires Slimb, which is thought to direct ubiquitin-mediated proteolysis of specifically phosphorylated forms of proteins, including Ci (ref. 18). There was much less Ci-155 in *slimb*² *Su(fu)* double-mutant clones (Fig. 5e) than in *slimb*² clones (Fig. 5a). Also, loss of *Su(fu)* reduced the accumulation of Ci-155 in *slimb*¹/*slimb*² wing discs, once again without increasing Ci-75 (Fig. 4). Thus, *Su(fu)* must affect Ci-155 levels by a mechanism other than the established Slimb-dependent conversion to Ci-75. The amount of *ci* RNA was not affected by *Su(fu)* mutation in whole discs (Fig. 5c, g). To account for the increased expression of Hh-target genes and the decrease in Ci-155 in *Su(fu)* mutants, we propose that the function of *Su(fu)* is to antagonize the conversion of Ci-155 to a more labile but transcriptionally active form in cells that are undergoing Hh signal transduction.

If *Su(fu)* blocks maturation of Ci-155 into an activator by binding to it⁷, then the concentration of Ci-155 relative to *Su(fu)* should affect the specific activity of Ci-155. Inactivation of one copy of the *ci* gene roughly halved the amount of Ci (Fig. 2n), but almost eliminated the effects of *PKA*-mutant clones on growth and patterning of wings. Normally, *PKA*-mutant clones are bordered on each side by an ectopic vein 3 and induce pattern duplications if located anywhere anterior to vein 3 (Fig. 1e). In *ci*/+ animals, *PKA*-mutant clones only induced additional anterior tissue if the clone formed anterior to vein 2 and in these cases induced an ectopic vein 3 in the middle of the clone (Fig. 1f). The normal organizing properties of *PKA*-mutant clones were restored in *ci* heterozygotes

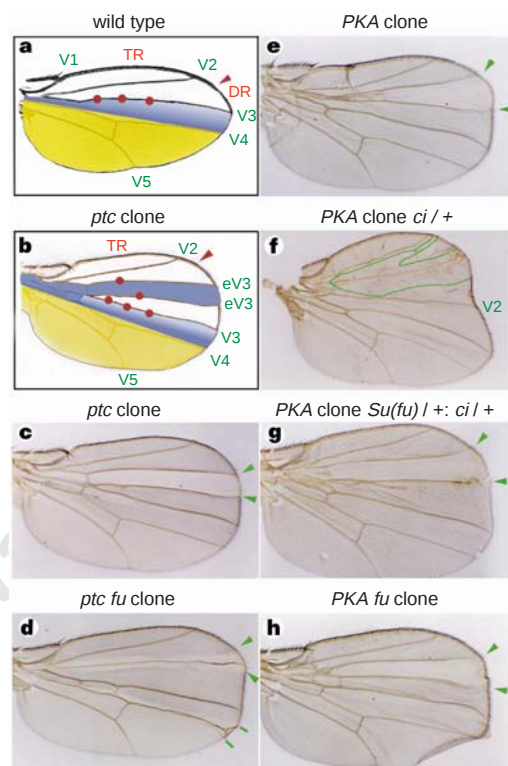


Figure 1 Effects of Ci/*Su(fu)* ratio and *Fu* kinase activity on wing patterning. **a**, Posterior compartment expression of Hh (yellow) and Hh-target genes, *dpp* (light blue) and *en* (dark blue), from third-instar wing discs projected on to a wild-type adult wing. At the edge of Hh signalling territory, vein 3 (V3), sensillae (red circles), and transition of marginal bristles to double row (DR) can be seen. **b**, Anterior *ptc* or *PKA* mutant clones between V3 and V2 induce ectopic Hh-target gene expression cell-autonomously (blue), ectopic vein 3 (eV3) and sensillae (red circles) either side of the clone, as well as anterior displacement of both V2 and the transition (arrowhead) to triple-row (TR) bristles. **c–i**, Mutant clones are seen between arrowheads and appear pale, except in **e**, where the clone is outlined in green. **c**, *ptc* clone; **d**, *ptc fu* double-mutant clone; **e**, *PKA* clone; **f**, *PKA* clone in *ci*/+ fly; **g**, *PKA* clone in *Su(fu)*/+; *ci*/+ fly; **h**, *PKA fu* double-mutant clone. **d**, A second *ptc fu* clone at the AP border (between the green lines) invaded the posterior, forming an ectopic vein 3 plus sensillae in place of vein 4, as did anterior *fu* clones at the AP border (not shown).

Table 1 *Su(fu)* and *Fu* modify wing phenotypes due to Hh signalling

		Distance to vein 2	Sensillae			Distance to vein 2	Sensillae
WT	F	41 ± 2	3.0	<i>Su(fu)</i>	F	46 ± 2	3.4
WT	M	40 ± 1	3.0	<i>Su(fu)</i>	M	45 ± 2	3.6
				<i>fu Su(fu)</i>	M	45 ± 2	3.2
<i>ptc</i> *	F	38 ± 1	0.5	<i>ptc Su(fu)</i> *	F	40 ± 2	2.1
<i>ptc</i>	M	38 ± 1	0.7				
<i>ptc fu</i> †	M	28 ± 3	0	<i>ptc fu Su(fu)</i> †	M	40 ± 2	2.6
<i>PKA</i> ‡	F	36 ± 2	0.1	<i>PKA Su(fu)</i> ‡	F	40 ± 2	2.0
<i>PKA</i>	M	33 ± 2	0				
<i>PKA fu</i>	M	32 ± 1	0.1	<i>PKA fu Su(fu)</i>	M	38 ± 3	2.3

Distances to vein2 from vein3 (WT) or anterior edge of mutant clone (*ptc*, *PKA*, *ptc fu*, *PKA fu*) are expressed as the sum of medial triple row and dorsal double-row bristles ± s.d. for females (F) and males (M). The number of sensillae distal to the vein 3–vein 4 crossvein that were on vein 3 (WT) or immediately anterior to mutant clones were counted. All values are means from at least 13 samples except for *ptc fu Su(fu)*.

* Also heterozygous for *fu* and *PKA*.

† Also heterozygous for *PKA*.

‡ Also heterozygous for *fu*.

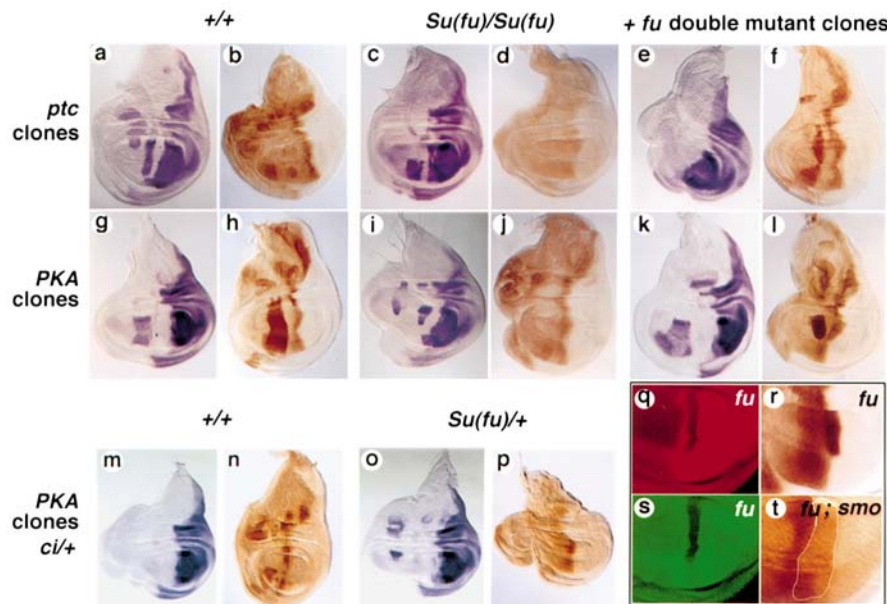


Figure 2 Effects of Ci/Su(fu) ratio and Fu kinase activity on anterior expression of En and Ci-155. **a–d**, *ptc* clones in discs with (+/+) or without *Su(fu)* activity; **g–j**, *PKA* clones in discs with (+/+) or without *Su(fu)* activity; **m–p**, *PKA* clones in *ci* heterozygotes that are wild-type (+/+) or heterozygous for *Su(fu)*; **e, f**, *ptc fu* double-mutant clones; **k, l**, *PKA fu* double-mutant clones; **q–s**, *fu* clones; **t**, *fu smo* double-mutant clone. En antibody staining (blue; red in **q**) was seen in the posterior (right) of all discs and at similar intensity in: **a**, *ptc*-mutant clones; and in **c**, *ptc*; or **i**, *PKA*-mutant clones in *Su(fu)/Su(fu)* animals. Less ectopic En was induced in: **g**, *PKA*-mutant clones; **k**, *PKA fu* double-mutant clones; and **o**, *PKA*-mutant clones in *Su(fu)/+*; *ci/+* animals. No ectopic En was induced in: **e**, *ptc fu* double-mutant clones; **m**, *PKA*-mutant clones in *ci/+* animals; or **q, fu** clones at the AP border. The presence of anterior clones in **e** and **m** is evident from disc distortions and ectopic Ci in sister discs (**f, n**). The *fu* clone in **q** is marked by the

absence of *arm-lacZ* staining (green) in **s**. **b**, Ci-155 stained with antibody 2A1 (brown) was exclusively anterior and was similarly increased at the AP border and in *ptc*-mutant clones. The highest ectopic expression of Ci was in: **h**, *PKA*- and **i**, *PKA fu*-mutant clones. There was more Ci than at the AP border in: **f**, *ptc fu* clones; and **r**, *fu* clones at the AP border, which projected into the posterior of the disc. **t**, *fu* clones also mutant for *smo* (outlined in white) projected into the posterior but stained only weakly for Ci-155 in the anterior, as seen for *smo* clones. In *Su(fu)/Su(fu)* animals, in **d** there was no increase in Ci-155 in *ptc* clones (disc distortion and ectopic En in sister discs (**c**) show the presence of clones); and in **j**, the amount of Ci-155 was only slightly raised in *PKA*-mutant clones. Ci staining in *PKA*-mutant clones was roughly halved in **n**, *ci/+* discs, and was even less in **p**, *Su(fu)/+*; *ci/+* discs.

by additional inactivation of one copy of the *Su(fu)* gene (Fig. 1g). Similarly, induction of En expression in *PKA*-mutant wing disc clones was lost in *ci* heterozygotes but was restored by loss of one copy of *Su(fu)* (Fig. 2g, m, o). Hence, the repressive effects of *Su(fu)* on Hh signal transduction depend on the relative concentrations of Ci and *Su(fu)*, consistent with a stoichiometric functional interaction.

The Fused protein appears to play a structural role in the cytoplasmic Ci-155 complex that helps to regulate proteolysis of Ci-155 to Ci-75 (ref. 8, J.T.O., unpublished results), but the function of its kinase activity has not been explained. A strong *fu* mutation, *fu*^{mH63}, which affects only the kinase domain, completely blocks transcriptional responses to Hh in the embryo but only partially in wing discs^{19,20}. Anterior wing-disc clones lacking both *ptc* and Fu kinase activity did not express ectopic En (Fig. 2e). The *ptc fu*^{mH63} clones induced ectopic sensillae and an ectopic vein 3 in the middle of the clone, and a much smaller displacement of vein 2 than *ptc* mutant clones (Fig. 1c, d; Table 1). Likewise, anterior *fu*^{mH63} clones at the AP border did not express En, invaded posterior compartment territory^{21,22} (Fig. 2q, s), and induced a central vein 3 decorated with sensillae at the position normally occupied by vein 4 (Fig. 1d; see legend).

Although it reduced the response to Hh and to *ptc* mutations in wing discs, the *fu*^{mH63} mutation increased the amount of Ci-155 at the AP border (Figs 2r; 5b, d) and in *ptc* mutant clones (Fig. 2b, f). Increased accumulation of Ci-155 was also evident in western blots of *fu*^{mH63} wing discs and was not accompanied by a decrease in Ci-75 (Fig. 4). Neither the level of Ci nor the wing morphology was affected by *fu*^{mH63} in mutant clones outside the Hh signalling region (not shown) or at the AP border when Hh signalling was blocked by

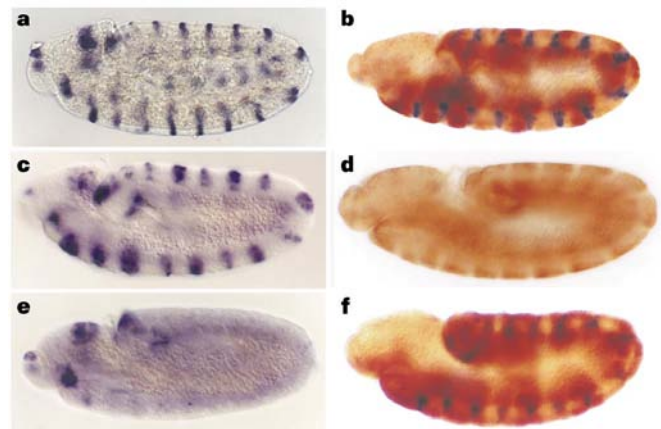


Figure 3 Loss of *Su(fu)* in embryos enhances ectopic *wg* expression and decreases Ci-155 levels resulting from PKA inhibition. **a, c, e**, *wg* RNA (blue); **b, d, f**, Ci-155 stained with antibody 2A1 (brown); in **b, f**, antibody staining for En (blue) is shown in stage-10-11 embryos. All embryos expressed the PKA-inhibitor transgene *UAS-R** under *prd-GAL4* control¹⁶. This led to: **a**, normal *wg* expression; and **b**, greatly increased Ci staining anterior to En stripes 1, 2, 3, 5, 7, 9, 11 and 13, where PKA is inhibited by *R** expression¹⁶. **c, d**, Embryos additionally lacking maternal and zygotic *Su(fu)* activity showed: **c**, expanded *wg* stripes 1, 2, 3, 5, 7, 9, 11 and 13, but **d**, no increase in Ci staining in segments where PKA was inhibited. **e, f**, Embryos additionally lacking maternal and zygotic Fu kinase activity: **e**, lost *wg* expression in trunk segments; and **f**, displayed increased Ci staining in regions of PKA inhibition. *fu* male embryos from germline clones were identified in **e**, by the absence of Sex-lethal antibody staining, and in **f**, according to decay of En expression.

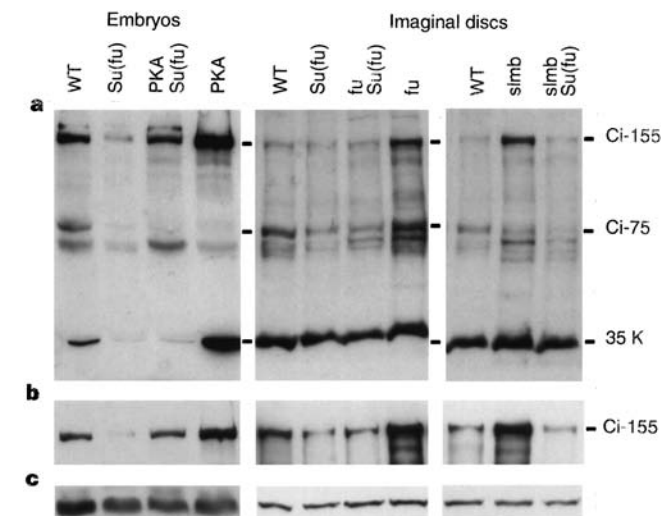


Figure 4 Western blots of embryo and wing disc extracts. **a**, Antibody raised against the zinc-finger domain of Ci recognizes Ci-155, Ci-75, a 35K band, and background bands smaller than Ci-75. **b**, Antibody 2A1 recognizes only Ci-155. **c**, Equal loading was determined by protein concentration ($\sim 20 \mu\text{g}$ protein per lane) and re-probing with antibody to the PKA type-II regulatory subunit (RII) for embryos, and according to disc number (25 discs per lane; 20 for *slimb* mutants) and re-probing with antibody to α -tubulin for wing discs. WT, wild type. In embryos, *PKA* denotes ubiquitous expression of *UAS-R** using *E22C-GAL4* (ref. 16) to inhibit PKA, and *Su(fu)* denotes loss of maternal and zygotic activity. In discs, *Su(fu)* and *fu* denote loss of zygotic *Su(fu)* and Fu kinase activity (*fu*^{mH63}). *slimb* denotes *slimb*¹/*slimb*² discs.

a null *smoothened* (*sno*) mutation^{23,24} (Fig. 2t). Hence, *fu*^{mH63} and *Su(fu)* mutations induced opposite effects on Hh-target gene expression and Ci levels that were in both cases limited to cells undergoing Hh signal transduction. In the absence of both *Su(fu)* and Fu kinase activity, the amount of Ci and the wing phenotype were typical of *Su(fu)* mutations²⁰ (Figs 4; 5h; Table 1). We suggest that Fu kinase normally opposes the action of *Su(fu)*, perhaps by promoting its dissociation from the Ci-155 complex.

This model predicts that Fu kinase activity is required only to activate Ci-155 that is associated with *Su(fu)*. The effects of Ci and *Su(fu)* dosage on *PKA*-mutant phenotypes in wing discs (Figs 1e–g; 2g, m, o) suggest that the amount of Ci-155 that is induced by *PKA* mutation overwhelms the inhibitory capacity of *Su(fu)*. If this excess Ci-155 is free of *Su(fu)*, it should be insensitive to Fu kinase. In agreement, we found that ectopic En and the wing phenotype induced by *PKA*-mutant clones was virtually unaffected by the *fu*^{mH63} mutation in *PKA fu*^{mH63} double-mutant clones (Figs 1e, h; 2g, k; Table 1). *ptc fu*^{mH63} double-mutant clones induced less Ci-155 than did *PKA fu*^{mH63} clones (Fig. 2f, l) but still retained some pattern-organizing activity in wings (Fig. 1d; Table 1), indicating that Ci-155 may be in excess relative to *Su(fu)*. By contrast, Fu kinase activity is essential for the induction of Hh-target genes in embryos, whether this is in response to Hh protein, *ptc* mutation⁵, or inhibition of PKA (Fig. 3e), even though PKA inhibition induced very high amounts of Ci-155 (Fig. 3f). We suggest that *Su(fu)* levels in embryos are comparatively high and are not exceeded by those of Ci-155 during Hh signalling or when PKA is inhibited.

Hh protein and *ptc* mutations elicit stronger phenotypic responses in embryos and wing discs than *PKA* mutations, despite the lower induction of Ci-155, indicating that Hh does more than simply inhibit the proteolysis of Ci-155 to Ci-75 (ref. 16). We have shown that the greater potency of Hh and *ptc* mutations relative to *PKA* mutations was suppressed or even reversed by eliminating *Su(fu)* or Fu kinase activity. We conclude that Hh promotes the Fu-kinase-dependent maturation of Ci into a transcriptional activator

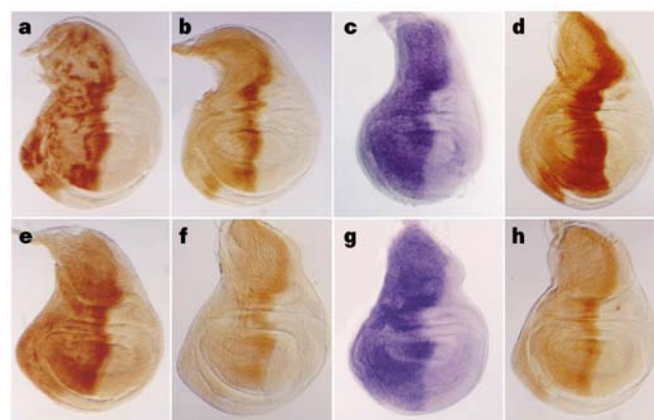


Figure 5 Effects of loss of *Su(fu)* on the level of Ci-155 are epistatic to *slimb* and *fu* mutations. **a, b, d, e, f, h**, Ci-155 staining with antibody 2A1 (brown): **a**, small *slimb*² anterior (left) clones induced high levels of Ci-155, whereas in **e**, Ci-155 was barely detectable above anterior levels in *slimb*² *Su(fu)* double-mutant clones; **b**, wild-type wing disc shows increased Ci staining at the AP border, which in **d** was increased in intensity and width in discs lacking Fu kinase activity (*fu*^{mH63}), and in **f** and **h**, was decreased in intensity in discs lacking *Su(fu)* (**f**) or both *Su(fu)* and Fu kinase activity (**h**). **c, g**, *ci* RNA (blue) levels in wild-type discs (**c**) were unaltered in the absence of *Su(fu)* (**g**).

that is otherwise antagonized by *Su(fu)*; this feature of Hh signalling is reproduced by *ptc* mutation but not by *PKA* mutation.

Hh signalling thus elicits several changes that are required to convert Ci into an effective transcriptional activator. Hh spares Ci-155 from *PKA*- and *Slimb*-dependent proteolysis to Ci-75 (refs 1, 18), perhaps by modifying the phosphorylation status of Ci (ref. 25), and promotes dissociation of the Ci-155 complex from microtubules⁸. We propose that in wing discs some of this 'primed' Ci-155 is not associated with *Su(fu)* and can activate *dpp* and *ptc* but not anterior *en* expression. Most of the primed Ci-155 in wing discs and perhaps all of the primed Ci-155 in embryos is inactive while it is in complex with *Su(fu)* and signalling by Hh and Fu kinase are necessary for it to become a transcriptional activator. This active form of Ci appears to be unstable and so is not detectable in the nuclei of cells responding to Hh.

The lower levels of Ci-155 that are found in wing discs close to the source of Hh compared with those in more distant regions of the Hh-signalling domain^{12,26} may be explained by our model if more Hh is required to stimulate the Fu-kinase-dependent step in Ci activation than to protect Ci-155 from proteolytic degradation to Ci-75. This dosage dependence may account for the restricted range of anterior *en* induction relative to *dpp* and *ptc* in wing discs^{11,12,26} and the single-cell range of Hh signalling in the embryonic ectoderm. □

Methods

Genetic analysis. Most wing disc clones were generated by FRT-mediated recombination in first instar larvae using the following alleles: *DCO*^{B3} (null allele of the principal PKA catalytic subunit)²⁷, *ptc*^{S2} (strong hypomorph), *fu*^{mH63} (strong hypomorphic class I allele)^{19,20}, *sno*^{11X43} (null allele)²⁴, *slimb*¹ (hypomorphic) and *slimb*² (probably null)¹⁸. *P[fru]* rescue transgenes²⁸ on 2L and 2R were used to generate *ptc fu*, *PKA fu* and *sno fu* double-mutant clones. *Su(fu)*^{LP} is null²⁹ and *Df(4)M62f* eliminates both *ci* and *pangolin* gene activity³⁰. **Analysis of wings.** Wings were mounted in Canada balsam/methyl salicylate (1:1) and clones marked by *stc* or *sha* and *y* at the margin were characterized for

each genotype only if they occupied congruent dorsal and ventral surfaces and extended from the wing margin up to or near the vein 3–4 crossvein.

Imaginal disc and embryo staining. Standard protocols were followed, using the anti-Ci antibody 2A1 and the anti-En antibody 4D9 with Vectastain Elite ABC kits and horseradish peroxidase staining for comparing the staining intensity of samples processed in parallel, or immunofluorescent secondary antibodies to confirm the location of clones.

Western blots. Approximately 50 μ l packed stage-11–12 embryos were homogenized in 150 μ l RIPA buffer before adding an equal volume of loading buffer to give ~ 0.4 mg ml⁻¹ protein. Generally, 100 wing discs were dissected into 100 μ l loading buffer and vortexed to homogenize them. Blots were probed successively with antibodies raised against the Ci zinc-finger domain (from P. W. Ingham), the Ci C-terminal half¹⁷, and the RII subunit of PKA or α -tubulin, and visualized using horseradish peroxidase-coupled secondary antibody and a chemiluminescent reagent (Pierce).

Received 14 September; accepted 27 October 1998.

1. Aza-Blanc, P., Ramirez-Weber, F.-A., Laget, M.-P., Schwarz, C. & Kornberg, T. B. Proteolysis that can be inhibited by Hedgehog directs Cubitus interruptus protein to the nucleus and changes its activity as a transcriptional regulator. *Cell* **89**, 1043–1053 (1997).
2. Ingham, P. W. Transducing Hedgehog: the story so far. *EMBO J.* **17**, 3505–2511 (1998).
3. Alexandre, C., Jacinto, A. & Ingham, P. W. Transcriptional activation of *hedgehog* target genes in *Drosophila* is mediated directly by the *Cubitus interruptus* protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes Dev.* **10**, 2003–2013 (1996).
4. Von Ohlen, T. & Hooper, J. E. Hedgehog signaling regulates transcription through Gli/Ci binding sites in the *wingless* enhancer. *Mech. Dev.* **68**, 149–156 (1997).
5. Forbes, A. J., Nakano, Y., Taylor, A. M. & Ingham, P. W. Genetic analysis of hedgehog signalling in the *Drosophila* embryo. *Development* (suppl.) **115**–124 (1993).
6. Akimaru, H. *et al.* *Drosophila* CBP is a co-activator of Cubitus interruptus in Hedgehog signalling. *Nature* **386**, 735–738 (1997).
7. Monnier, V., Dussillol, F., Alves, G., Lamour-Isnard, C. & Plessis, A. Suppressor of fused links Fused and Cubitus interruptus on the Hedgehog signalling pathway. *Curr. Biol.* **8**, 583–586 (1998).
8. Robbins, D. J. *et al.* Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein Costal2. *Cell* **90**, 225–234 (1997).
9. Sisson, J. C., Ho, K. S., Suyama, K. & Scott, M. P. Costal2, a novel kinesin-related protein in the hedgehog signaling pathway. *Cell* **90**, 235–245 (1997).
10. Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by Hedgehog protein. *Nature* **368**, 208–214 (1994).
11. Blair, S. *engrailed* expression in the anterior lineage compartment of the developing wing blade of *Drosophila*. *Development* **115**, 21–33 (1992).
12. Strigini, M. & Cohen, S. M. A Hedgehog activity gradient contributes to AP axial patterning of the *Drosophila* wing. *Development* **124**, 4697–4705 (1997).
13. Sturtevant, M. A., Biehs, B., Marin, E. & Bier, E. The *spalt* gene links the A/P compartment boundary to a linear adult structure in the *Drosophila* wing. *Development* **124**, 21–32 (1997).
14. Hidalgo, A. Three distinct roles for the *engrailed* gene in *Drosophila* wing development. *Curr. Biol.* **4**, 1087–1098 (1994).
15. Ingham, P. W. Localized *hedgehog* activity controls spatial limits of *wingless* transcription in the *Drosophila* embryo. *Nature* **366**, 560–562 (1993).
16. Ohlmeyer, J. T. & Kalderon, D. Dual pathways for induction of *wingless* by protein kinase A and Hedgehog in *Drosophila* embryos. *Genes Dev.* **11**, 2250–2258 (1997).
17. Motzny, C. K. & Holmgren, R. The *Drosophila* Cubitus interruptus protein and its role in the *wingless* and *hedgehog* signal transduction pathways. *Mech. Dev.* **52**, 137–150 (1995).
18. Jiang, J. & Struhl, G. Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* **391**, 493–496 (1998).
19. Busson, D., Limbourg-Bouchon, B., Mariol, M.-C., Preat, T. & Lamour-Isnard, C. Genetic analysis of viable and lethal *fused* mutants of *Drosophila melanogaster*. *Roux Arch. Dev. Biol.* **197**, 221–230 (1988).
20. Preat, T. *et al.* Segmental polarity in *Drosophila melanogaster*: genetic dissection of *fused* in a *Suppressor of fused* background reveals interaction with *costal-2*. *Genetics* **135**, 1047–1062 (1993).
21. Rodriguez, I. & Basler, K. Control of compartmental affinity boundaries by Hedgehog. *Nature* **389**, 614–617 (1997).
22. Blair, S. S. & Ralston, A. Smoothed-mediated Hedgehog signalling is required for the maintenance of the anterior–posterior lineage restriction in the developing wing of *Drosophila*. *Development* **124**, 4053–4063 (1997).
23. Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M. & Hooper, J. E. The *Drosophila* *smoothed* gene encodes a seven-pass membrane protein, a putative receptor for the Hedgehog signal. *Cell* **86**, 221–232 (1996).
24. van den Heuvel, M. & Ingham, P. W. *smoothed* encodes a receptor-like serpentine protein required for Hedgehog signalling. *Nature* **382**, 547–551 (1996).
25. Chen, Y., Gallaher, N., Goodman, R. H. & Smolik, S. M. Protein kinase A directly regulates the activity and proteolysis of Cubitus interruptus. *Proc. Natl Acad. Sci. USA* **95**, 2349–2354 (1998).
26. Mullor, J. L., Calleja, M., Capdevila, J. & Guerrero, I. Hedgehog activity, independent of Decapentaplegic, participates in wing disc patterning. *Development* **124**, 1227–1237 (1997).
27. Lane, M. E. & Kalderon, D. Genetic investigation of cAMP-dependent protein kinase function in *Drosophila* development. *Genes Dev.* **7**, 1229–1243 (1993).
28. Therond, P. *et al.* Molecular organization and expression pattern of the segment polarity gene *fused* of *Drosophila melanogaster*. *Mech. Dev.* **44**, 165–180 (1993).
29. Preat, T. Characterization of *Suppressor of fused*, a complete suppressor of the *fused* segment polarity gene of *Drosophila melanogaster*. *Genetics* **132**, 725–736 (1992).
30. Brunner, E., Peter, O., Schweizer, L. & Basler, K. *pangolin* encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in *Drosophila*. *Nature* **385**, 829–833 (1997).

Acknowledgements. We thank R. Holmgren, P. Ingham, J. Jiang, T. Kornberg, C. Lamour-Isnard, N. Patel, N. Perrimon and A. Plessis for reagents; N. Baker, J. Erickson and J. Mohler for discussions, and the NIH for supporting this work.

Correspondence and requests for materials should be addressed to D.K. (e-mail: ddk1@columbia.edu).

Local calcium signalling by inositol-1,4,5-trisphosphate in Purkinje cell dendrites

Elizabeth A. Finch & George J. Augustine

Department of Neurobiology, Duke University Medical Center, PO Box 3209, Durham, North Carolina 27710, USA

The second messenger inositol-1,4,5-trisphosphate (InsP₃) releases Ca²⁺ from intracellular Ca²⁺ stores by activating specific receptors on the membranes of these stores¹. In many cells, InsP₃ is a global signalling molecule that liberates Ca²⁺ throughout the cytoplasm^{1,2}. However, in neurons the situation might be different^{3,4}, because synaptic activity may produce InsP₃ at discrete locations. Here we characterize InsP₃ signalling in

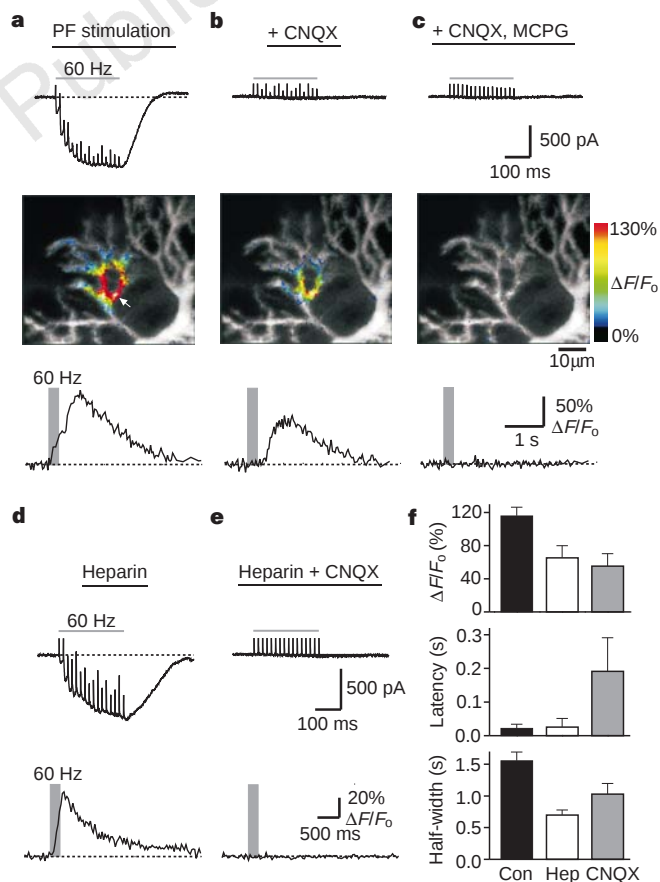


Figure 1 Components of postsynaptic Ca²⁺ signalling during parallel-fibre (PF) synaptic activity. **a–c**, Purkinje cell responses to repetitive PF stimulation (250 ms, 60 Hz; indicated by bars or shaded boxes) in **a**, control conditions, **b**, CNQX (15 μ M), and **c**, CNQX (15 μ M) plus MCPG (1 mM). Top, electrical responses. Middle, Ca²⁺ signals (pseudocolour), measured at the peak of the response, superimposed on a reconstructed image of the Purkinje cell (greyscale). Bottom, time course of Ca²⁺ signals, measured in a 5- μ m dendritic segment (arrow in **a**). **d**, **e**, Responses to repetitive PF stimulation (250 ms, 60 Hz) in Purkinje cells dialysed with heparin (50 μ g ml⁻¹) in **d**, the absence or **e**, the presence of CNQX (15 μ M). Top, electrical responses. Bottom, time course of the Ca²⁺ signal. **f**, Properties of Ca²⁺ signals produced by PF stimulation (250 ms, 60 Hz) in control conditions ($n = 4$), in cells dialysed with heparin ($n = 3$), and in cells treated with CNQX ($n = 4$).

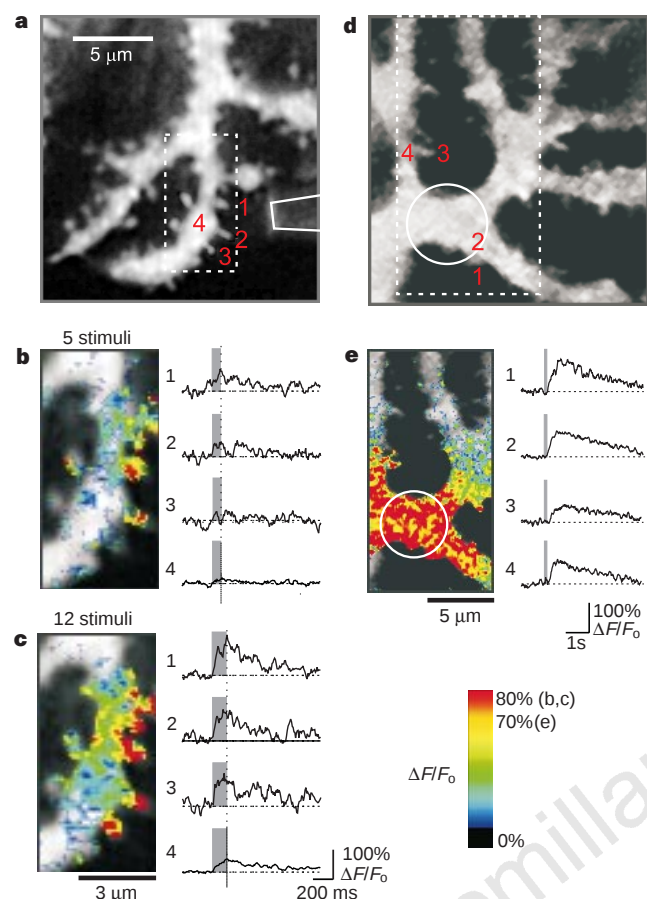


Figure 2 InsP_3 -mediated Ca^{2+} signals produced in Purkinje cell dendritic spines and shafts by parallel-fibre (PF) activity (left) and uncaging InsP_3 (right). **a**, The area studied in **b**, **c** (outlined by dashed box) and the position of the PF stimulation pipette are shown. Numbers indicate locations from which Ca^{2+} signals were measured in **b**, **c**. **b**, Ca^{2+} signals produced by 5 stimuli (50 Hz) in CNQX. Traces represent Ca^{2+} signals in the numbered regions of **a**, **c**. **c**, Ca^{2+} signals produced by 12 stimuli (80 Hz). **d**, Circle indicates position of the ultraviolet light spot (5 μm diameter) and numbers indicate locations from which Ca^{2+} signals were measured in **e**. **e**, Ca^{2+} signals produced by uncaging InsP_3 .

postsynaptic cerebellar Purkinje neurons, which have a high level of InsP_3 receptors⁵. We find that repetitive activation of the synapse between parallel fibres and Purkinje cells causes InsP_3 -mediated Ca^{2+} release in the Purkinje cells. This Ca^{2+} release is restricted to individual postsynaptic spines, where both metabotropic glutamate receptors^{6,7} and InsP_3 receptors⁵ are located, or to multiple spines and adjacent dendritic shafts. Focal photolysis of caged InsP_3 (ref. 8) in Purkinje cell dendrites also produces Ca^{2+} signals that spread only a few micrometres from the site of InsP_3 production. Uncaged InsP_3 produces a long-lasting depression of parallel-fibre synaptic transmission that is limited to synapses where the Ca^{2+} concentration is raised. Thus, in Purkinje cells InsP_3 acts within a restricted spatial range that allows it to regulate the function of local groups of parallel-fibre synapses.

To determine whether synaptic activity evokes InsP_3 -mediated Ca^{2+} release, we stimulated parallel fibres while monitoring electrical responses with whole-cell patch-clamp recording and postsynaptic Ca^{2+} levels with confocal microscopy in individual Purkinje cells. Repetitive activation of parallel-fibre synapses produced a transient and local rise in postsynaptic Ca^{2+} concentration^{9–12} (Fig. 1a). These Ca^{2+} signals consisted of an initial component that coincided with excitatory postsynaptic currents (EPSCs) and a second, delayed, component that often peaked after synaptic electrical activity ended. 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX), an antagonist of AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate)-type glutamate receptors, completely blocked both the EPSCs¹³ (Fig. 1b, top) and the initial rise in Ca^{2+} , but did not block the delayed component of the Ca^{2+} signals (Fig. 1b; $n = 7$). The initial component thus appears to arise from Ca^{2+} influx through voltage-gated Ca^{2+} channels that are activated when AMPA receptors depolarize the postsynaptic dendrites⁹. RS - α -methyl-4-carboxyphenylglycine (MCPG; 1 mM), an antagonist of metabotropic glutamate receptors (mGluRs)⁷, blocked the delayed Ca^{2+} responses measured in the presence of CNQX ($97\% \pm 3.4\%$ reduction (mean $\pm$ s.e.m.); $n = 4$), showing that this component arises from activation of mGluRs (Fig. 1c). We also blocked this delayed component by dialysing Purkinje cells with heparin (50 $\mu\text{g ml}^{-1}$), which prevents InsP_3 -mediated Ca^{2+} release¹⁴ (Fig. 1d, e). These results show that glutamate released from parallel-fibre presynaptic terminals activates two Ca^{2+} signalling pathways in postsynaptic Purkinje cells: an initial, rapid influx of

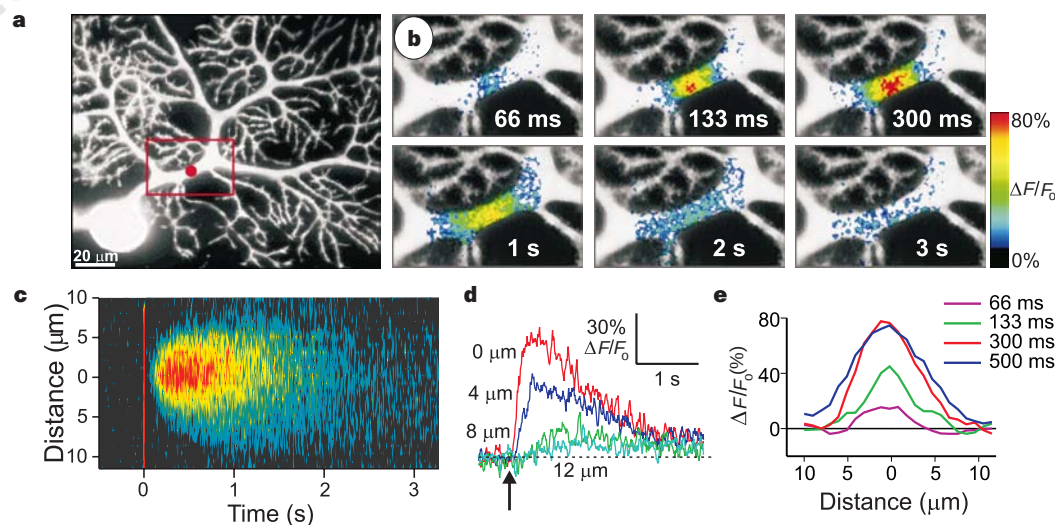


Figure 3 Spatiotemporal properties of dendritic Ca^{2+} signals produced by local photolysis of caged InsP_3 . **a**, Image of a Purkinje cell, with box indicating the area in **b** and circle illustrating the ultraviolet light spot. **b**, Images of the Ca^{2+} signal produced by InsP_3 at the indicated times after photolysis. **c**, Spatial range and time course of Ca^{2+} signal shown in **b**. The fluorescence-intensity profile was

measured along the length of the dendrite within each image during the response. 0 μm indicates the centre of the light spot and the red line the time of uncaging. **d**, Time courses of Ca^{2+} signals at the indicated distances from the ultraviolet light spot. Time of light flash indicated by arrow. **e**, Spatial extent of Ca^{2+} signal at indicated times after photolysis, measured along the dendrite as in **c**.

Ca^{2+} initiated by AMPA receptors and a delayed release of intracellular Ca^{2+} resulting from production of InsP_3 by mGluRs.

We isolated the two components of local Ca^{2+} signalling evoked by parallel-fibre activity by examining Ca^{2+} influx in the presence of heparin and InsP_3 -mediated Ca^{2+} release in the presence of CNQX. Postsynaptic Ca^{2+} release by InsP_3 was not detected following single stimuli to parallel fibres, but was observed following as few as five stimuli (Fig. 2a). For 15 stimuli at 60 Hz, roughly half of the Ca^{2+} signal amplitude resulted from InsP_3 -mediated Ca^{2+} release and half was due to Ca^{2+} influx (Fig. 1f, top). The average latency for InsP_3 -mediated Ca^{2+} release was 190 ms, about ten times longer than for Ca^{2+} influx (Fig. 1f, middle), and blocking Ca^{2+} release reduced the duration of parallel-fibre-mediated Ca^{2+} signals by 55% (Fig. 1f, bottom). InsP_3 -mediated Ca^{2+} signals were more spatially restricted than those observed in the absence of CNQX; for 15 stimuli at 60 Hz, the spatial extent of the InsP_3 -mediated Ca^{2+} -release signal was only $37\% \pm 10\%$ ($n = 4$) of the combined Ca^{2+} influx and release signal (Fig. 1a, b). Thus, the delayed release of Ca^{2+} by InsP_3 determines the magnitude, duration and spatial range of Ca^{2+} signals evoked by parallel-fibre activity.

To study postsynaptic InsP_3 -mediated Ca^{2+} signals at the level of individual synapses (Fig. 2a), we decreased the stimulus intensity to activate fewer than ten parallel fibres¹⁵. A train of five parallel-fibre stimuli produced InsP_3 -mediated Ca^{2+} signals in Purkinje cells (Fig. 2b) that were restricted to individual postsynaptic spines (traces 1–3), whereas higher levels of activity increased the Ca^{2+} concentration (Fig. 2c) both within spines (traces 1–3) and, to a lesser degree, in the adjacent dendritic shaft (trace 4) and neighbouring spines. These results indicate that parallel-fibre activity generates InsP_3 within postsynaptic spines and InsP_3 may then diffuse to activate Ca^{2+} release in the adjacent dendritic shaft when sufficient levels of InsP_3 are produced.

To examine the precise spatial range of InsP_3 signalling, we photolysed caged InsP_3 in defined areas of 3–5 μm diameter (Fig. 2d). InsP_3 increased Ca^{2+} levels in dendritic shafts and spines at the uncaging site (Fig. 2e, traces 1, 2) and also in more distant dendrites and spines (traces 3, 4), showing that signalling by InsP_3 spreads along the dendrites and into spines. Ca^{2+} signals produced by InsP_3 appeared initially at the uncaging site (Fig. 3a–c) and were much smaller and slower only a few micrometres away from the site of InsP_3 production (Fig. 3d). In unbranched regions of dendrites, 10–20 μM InsP_3 produced a Ca^{2+} signal that reached a half-width of 9 μm (Fig. 3c, e). Lower InsP_3 levels produced more restricted Ca^{2+} signals, whereas maximal InsP_3 concentrations (80–160 μM) produced Ca^{2+} signals with an average half-width of $12.3 \pm 1.2 \mu\text{m}$ ($n = 10$). This restricted spatial range of InsP_3 action in Purkinje cell dendrites indicates that diffusing InsP_3 does not activate regenerative Ca^{2+} release very far beyond the site of InsP_3 production. The kinetics and amplitude of Ca^{2+} signals were very sensitive to InsP_3 concentration^{14,16,17}. The Ca^{2+} signal produced by 10–20 μM InsP_3 appeared after a short latency (<33 ms), peaked at 175 ms, and declined to basal levels within a few seconds (Fig. 3c, d). Near-threshold levels of InsP_3 (5–10 μM) produced smaller Ca^{2+} increases with a longer latency and slower rise time, whereas maximal concentrations of InsP_3 (80–160 μM) produced Ca^{2+} signals with a shorter latency, faster rise time and larger amplitude. These Ca^{2+} signals were in the micromolar range, as reported by the indicator Ca^{2+} Green-5N ($K_d = 14 \mu\text{M}$) and the activation of dendritic Ca^{2+} -activated K^+ channels (data not shown)¹⁸.

Comparing Ca^{2+} -release signals produced by stimulation of parallel fibres and uncaging of InsP_3 allowed us to estimate the amount of InsP_3 produced by synaptic activity. Although the latencies of the Ca^{2+} signals were not comparable, presumably because of delays associated with synaptic production of InsP_3 , the amplitude and rate of rise of the Ca^{2+} signals produced by 16 stimuli to parallel fibres (Fig. 1b) were roughly equivalent to those produced by 10–20 μM InsP_3 (Fig. 3d). This indicates that each

parallel-fibre stimulus produces of the order of 1 μM InsP_3 and that Ca^{2+} release requires repetitive synaptic activity, because a single response produces a subthreshold amount of InsP_3 .

One postulated^{19–22} physiological role for postsynaptic InsP_3 is in the long-term depression (LTD)^{23,24} of parallel-fibre synapses. LTD is a form of plasticity that is spatially restricted^{23–25} and requires both increases in dendritic Ca^{2+} levels and activation of mGluRs^{23,24}. We determined whether local postsynaptic Ca^{2+} signalling by InsP_3 is sufficient to induce LTD by repeatedly (with 50 pulses, 1 Hz) photoreleasing high levels of InsP_3 (80–160 μM) at sites at which parallel-fibre synapses were activated (Fig. 4a). This produced spatially restricted Ca^{2+} -release signals (Fig. 4a, right) that were sustained throughout the train of stimuli, yielding multiple Ca^{2+} transients superimposed on an elevated plateau (Fig. 4b). Such Ca^{2+} signals produced a depression of parallel-fibre synaptic transmission ($26 \pm 3\%$ at 60 min; Fig. 4c, d) that lasted for the duration of the recording. InsP_3 -induced LTD did not require parallel-fibre activity during uncaging, showing that InsP_3 alone is sufficient to produce LTD. However, large and persistent Ca^{2+} signals were required; we observed LTD in 5 out of 6 Purkinje cells in which

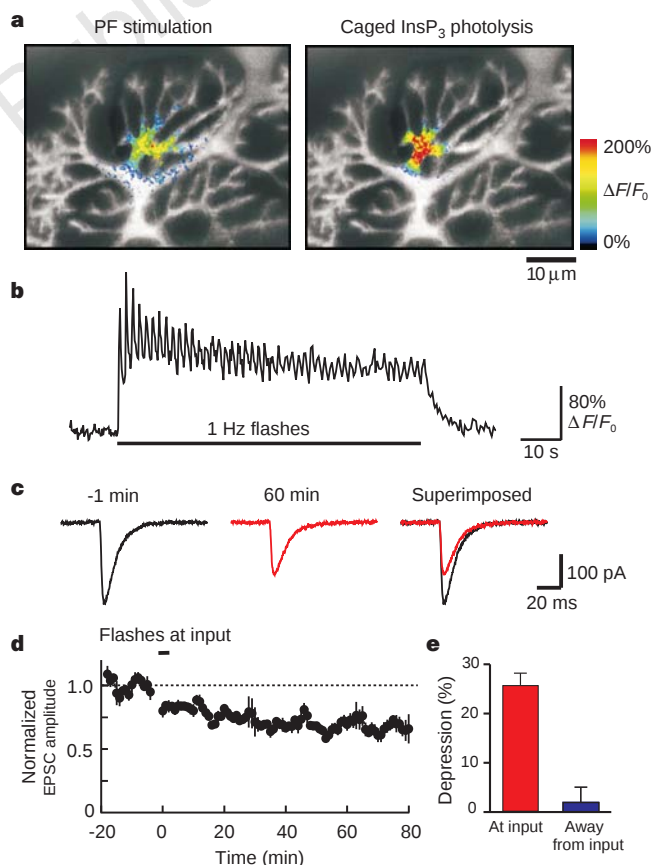


Figure 4 InsP_3 production induces a spatially restricted long-term depression of parallel-fibre (PF) synaptic transmission. **a**, Left, active PF synapses were located by monitoring the dendritic Ca^{2+} signal produced by a brief train of PF stimuli (230 ms, 40 Hz). Right, a train of ultraviolet light flashes (50 flashes, 1 Hz) was then delivered at this location; the second flash produced the Ca^{2+} signal illustrated. **b**, Time course of Ca^{2+} signal produced by repetitive photolysis of caged InsP_3 (50 flashes, 1 Hz). Measurement is from a different cell than to that shown in **a**. **c**, Averages of six EPSCs before and 60 min after the ultraviolet light flashes. **d**, Time course of changes in EPSC amplitude produced by repetitively photolysing caged InsP_3 (bar) at the active PF input ($n = 5$). EPSC amplitudes are normalized to the mean value before photolysing caged InsP_3 and data points are averages of six EPSCs (at 0.1 Hz). **e**, Mean amounts of LTD, measured either at 60 min or at the end of the experiment, produced by uncaging InsP_3 either at the synaptic input ($n = 5$) or 40–70 μm farther along the dendrite ($n = 3$).

InsP₃ produced large (>100%) and sustained Ca²⁺ signals, whereas synaptic transmission was not depressed when InsP₃ evoked a Ca²⁺ increase of less than 100% (*n* = 2) or that lasted less than 10 s (*n* = 2). The LTD of the parallel-fibre synaptic response was restricted to synapses near the site of InsP₃ production; large InsP₃-mediated Ca²⁺ signals generated 40–70 μm away from the active synapses produced no decline in EPSCs (Fig. 4d, e). Thus, InsP₃ is sufficient to produce a spatially restricted and long-lasting depression of parallel-fibre synaptic transmission that coincides with, and apparently results from, the highly localized intradendritic release of Ca²⁺.

Our results show that InsP₃ acts as a local postsynaptic signalling molecule that creates functional postsynaptic compartments in the spines and dendrites of cerebellar Purkinje cells. Such chemical signalling by InsP₃ can extend beyond single synapses (Fig. 3c) to influence other nearby synapses, but is much more restricted than electrical signalling, which can spread extensively through the dendritic arbor of the Purkinje cell. Limited spread of InsP₃ signalling is likely to result both from the properties of InsP₃ receptors—such as the nonlinearity of InsP₃-receptor activation^{14,16}, due in part to their regulation by Ca²⁺ (refs 14, 16), and the apparent low affinity¹⁷ of these receptors for InsP₃—and from other unique properties of Purkinje cells, such as efficient mechanisms for buffering and removing cytosolic Ca²⁺ (ref. 26), potential buffering of InsP₃ by a high density of InsP₃ receptors⁵, a possibly short lifetime for InsP₃ (ref. 27) and a large, highly branched dendritic arbor. Our observations extend other studies proposing that InsP₃ is involved in LTD^{20–22}. We have shown, first, that parallel-fibre activity generates InsP₃ and InsP₃-induced Ca²⁺ release, second, that InsP₃ signalling can spread beyond single synapses to nearby dendritic regions in a spatially limited manner, and third, that local InsP₃-mediated Ca²⁺ signals are sufficient to induce a spatially restricted LTD. Our data indicate that the restricted spread of LTD induced by pairing parallel-fibre and climbing-fibre activity²⁵ may arise from the local amplification of Ca²⁺ signals by InsP₃ produced at parallel-fibre synapses. Given that other forms of synaptic plasticity have a limited spatial spread and may involve mGluRs²⁸, local postsynaptic signalling by InsP₃ may be a general mechanism for encoding spatial information in neuronal dendrites. □

Methods

Patch-clamp recording. Sagittal cerebellar slices (200 μm) from 14–21-day-old rats²⁹ were perfused with oxygenated saline containing (in mM): 125 NaCl, 2.5 KCl, 2.5 CaCl₂, 26.2 NaHCO₃, 1.3 MgSO₄, 1.0 NaH₂PO₄, 20 D-glucose, 0.01–0.02 bicuculline methiodide (Sigma) and 0.05–0.1 Trolox ((±)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (in DMSO; Fluka)) to minimize phototoxicity. CNQX (Sigma) and MCPG (ToCris) were added as indicated. Whole-cell patch-clamp recordings were made using pipettes (1.5–2.5 MΩ) filled with internal solution containing (in mM): 130 potassium gluconate, 2 NaCl, 4 MgCl₂, 4 Na₂-ATP, 0.4 Na-GTP, 10 sodium phosphocreatine, 20 HEPES, pH 7.3, and either EGTA or Ca²⁺ indicator (100–250 μM). The relatively low Ca²⁺-chelator concentrations used should have had little impact on Purkinje cell Ca²⁺ buffering²⁶. Caged InsP₃ (ref. 30; 400–600 μM) and heparin (Sigma) were added as indicated. Parallel fibres were stimulated with a glass stimulating electrode (tip diameter, 2–4 μm) positioned in the molecular layer. Cells were voltage-clamped at –60 mV and parallel-fibre EPSCs were acquired and analysed using PCLAMP software (Axon Instruments). We delivered 5-mV hyperpolarizing test pulses to measure series and input resistance¹³. LTD experiments were accepted if the series resistance changed by <20%, the input resistance was greater than 150 MΩ, the holding current changed by <5% and Ca²⁺ levels returned to baseline following photolysis of caged InsP₃. Uncaging InsP₃ did not change the input resistance or holding current, and ultraviolet illumination in the absence of caged InsP₃ did not affect parallel-fibre synaptic transmission.

Optical techniques. Purkinje cells were loaded for at least 30 min with Ca²⁺ indicator (Calcium Green-1 or Oregon Green BAPTA-1, Molecular Probes; 100–250 μM) through the patch pipette. Ca²⁺ was measured^{9,12} using a high-

speed confocal microscope (Noran Odyssey XL) to acquire fluorescence images (30–240 Hz). Before Ca²⁺ levels were measured, depolarizing pulses were applied to activate voltage-gated Ca²⁺ channels to load Ca²⁺ stores. Ca²⁺ signals were measured as the change in fluorescence relative to baseline fluorescence ($\Delta F/F_0$). Resulting averages of eight images were thresholded at 10%, masked using a highly averaged image of the cell, encoded with a pseudocolour look-up table, and overlaid on a greyscale image of the Purkinje cell. Greyscale images were reconstructed from a series of confocal optical sections of the Purkinje cell acquired at the end of the experiment. Spread was measured by fitting the spatial profile of the Ca²⁺ signal with a gaussian function and measuring its width at half-maximal amplitude. The amount of caged InsP₃ photolysed in a 3–5-μm spot⁸ was varied by changing the intensity and/or duration of the pulse of ultraviolet light. The intensity of ultraviolet light was estimated by measuring the fluorescence of uncaged fluorescein dextran. InsP₃ production was then calculated by assuming that caged InsP₃ is photolysed at 54% of the efficiency of the caged fluorescein, on the basis of the extinction coefficients and quantum yields for these compounds³⁰.

Received 7 September; accepted 2 November 1998.

- Berridge, M. J. Inositol trisphosphate and cell signalling. *Nature* **261**, 315–325 (1993).
- Berridge, M. J. Elementary and global aspects of calcium signalling. *J. Physiol. (Lond.)* **499**, 291–306 (1997).
- Berridge, M. J. Neuronal calcium signaling. *Neuron* **21**, 13–26 (1998).
- Jaffe, D. B. & Brown, T. E. Metabotropic glutamate receptor activation induces calcium waves within hippocampal dendrites. *J. Neurophysiol.* **72**, 471–474 (1994).
- Walton, P. D. et al. F- and inositol trisphosphate receptors coexist in avian cerebellar Purkinje neurons. *J. Cell Biol.* **113**, 1145–1157 (1991).
- Nusser, Z., Mulvihill, E., Streit, P. & Somogyi, P. Subsynaptic segregation of metabotropic and ionotropic glutamate receptors as revealed by immunogold localization. *Neuroscience* **61**, 421–427 (1994).
- Batchelor, A. M., Madge, D. J. & Garthwaite, J. Synaptic activation of metabotropic glutamate receptors in the parallel fiber-Purkinje cell pathway in rat cerebellar slices. *Neuroscience* **63**, 911–915 (1994).
- Wang, S. S.-H. & Augustine, G. J. Confocal imaging and local photolysis of caged compounds: dual probes of synaptic function. *Neuron* **15**, 755–760 (1995).
- Eilers, J., Augustine, G. J. & Konnerth, A. Subthreshold synaptic Ca²⁺ signalling in fine dendrites and spines of cerebellar Purkinje neurons. *Nature* **373**, 155–158 (1995).
- Denk, W., Sugimori, M. & Llinas, R. Two types of calcium response limited to single spines in cerebellar Purkinje cells. *Proc. Natl Acad. Sci. USA* **92**, 8279–8282 (1995).
- Hartell, N. A. Strong activation of parallel fibers produces localized calcium transients and a form of LTD that spreads to distant synapses. *Neuron* **16**, 601–610 (1996).
- Eilers, J., Takechi, H., Finch, E. A., Augustine, G. J. & Konnerth, A. Local dendritic Ca²⁺ signaling induces cerebellar long-term depression. *Learning Memory* **3**, 159–168 (1997).
- Llano, I., Marty, A., Armstrong, C. M. & Konnerth, A. Synaptic- and agonist-induced excitatory currents of cerebellar Purkinje cells in rat cerebellar slices. *J. Physiol. (Lond.)* **434**, 183–213 (1991).
- Finch, E. A., Turner, T. J. & Goldin, S. M. Calcium as a coagonist of inositol trisphosphate-induced calcium release. *Science* **252**, 443–446 (1991).
- Barbour, B. Synaptic currents evoked in Purkinje cells by stimulating individual granule cells. *Neuron* **11**, 759–769 (1993).
- Parker, I., Yao, Y. & Ilyin, V. Fast kinetics of calcium liberation induced in *Xenopus* oocytes by photoreleased inositol trisphosphate. *Biophys. J.* **70**, 222–237 (1996).
- Khodakhah, K. & Ogden, D. Fast activation and inactivation of inositol trisphosphate-evoked Ca²⁺ release in rat cerebellar Purkinje neurons. *J. Physiol. (Lond.)* **487**, 343–358 (1995).
- DiChiara, T. J. & Reinhart, P. H. Distinct effects of Ca²⁺ and voltage on the activation and deactivation of cloned Ca²⁺-activated K⁺ channels. *J. Physiol. (Lond.)* **489**, 403–418 (1995).
- Berridge, M. J. A tale of two messengers. *Nature* **365**, 388–389 (1993).
- Kasano, K. & Hirano, T. Involvement of inositol trisphosphate in cerebellar long-term depression. *Neuroreport* **6**, 569–572 (1995).
- Khodakhah, K. & Armstrong, C. M. Induction of long-term depression and rebound potentiation by inositol trisphosphate in cerebellar Purkinje neurons. *Proc. Natl Acad. Sci. USA* **94**, 14009–14014 (1997).
- Inoue, T., Kato, K., Khoda, K. & Mikoshiba, K. Type 1 inositol 1,4,5-trisphosphate receptor is required for induction of long-term depression in cerebellar Purkinje neurons. *J. Neurosci.* **18**, 5366–5373 (1998).
- Ito, M. Long-term depression. *Annu. Rev. Neurosci.* **12**, 85–102 (1989).
- Linden, D. J. & Connor, J. A. Long-term synaptic depression. *Annu. Rev. Neurosci.* **18**, 319–357 (1995).
- Wang, S. S.-H. & Augustine, G. J. Lateral spread of postsynaptic plasticity from active to inactive cerebellar synapses. *Proc. Natl Acad. Sci. USA* (submitted).
- Fierro, L. & Llano, I. High endogenous calcium buffering in Purkinje cells from rat cerebellar slices. *J. Physiol. (Lond.)* **496**, 617–625 (1996).
- Yamada, M. et al. Ultrastructural localization of inositol 1,4,5-trisphosphate 3-kinase in rat cerebellar cortex. *Brain Res.* **578**, 41–48 (1992).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Doroshenko, P. A., Wopmann, A., Miljanich, G. & Augustine, G. J. Pharmacologically distinct presynaptic calcium channels in cerebellar excitatory and inhibitory synapses. *Neuropharmacol.* **36**, 865–872 (1997).
- Walker, J. W., Feeney, J. & Trentham, D. R. Photolabile precursors of inositol phosphates. Preparation and properties of 1-(2-nitrophenyl)ethyl esters of myo-inositol 1,4,5-trisphosphate. *Biochemistry* **28**, 3272–3280 (1989).

Acknowledgements. We thank J. Walker and J. Horne for synthesis of caged InsP₃; T. Blanpied, J. Kauer, F. Schweizer and S. Wang for helpful discussions; and J. Rongero for developing image-processing software. This was supported by the NIH.

Correspondence and requests for materials should be addressed to G.J.A. (e-mail: georgea@neuro.duke.edu).

A new class of synaptic response involving calcium release in dendritic spines

Hajime Takechi, Jens Eilers & Arthur Konnerth

I. Physiologisches Institut, Universität des Saarlandes, 66421 Homburg, Germany

In the classical view, transmission of signals across synapses in the mammalian brain involves changes in the membrane potential of the postsynaptic cell. The use of high-resolution cellular imaging has revealed excitatory synapses at which postsynaptic, transient alterations in calcium ion concentration are tightly associated with electrical responses (reviewed in ref. 1). Here, by investigating the synapse between parallel glutamatergic fibres and Purkinje cells in the mouse cerebellum, we identify a class of postsynaptic responses that consist of transient increases in dendritic Ca^{2+} concentration but not changes in somatic membrane potential. Our results indicate that these synaptic Ca^{2+} transients are mediated by activation of metabotropic glutamate-responsive mGluR1-type receptors²⁻⁴ and require inositol-1,4,5-trisphosphate-mediated Ca^{2+} release^{5,6} from intradendritic stores. The new type of synaptic response is restricted

to postsynaptic microdomains, which range, depending on the frequency of stimulation, from individual spines to small spinodendritic compartments. Thus, the synaptic Ca^{2+} -release signal may be one of the critical cues that determine the input specificity of long-term depression, a well-established form of activity-dependent plasticity at these synapses⁷⁻⁹.

The dendrites of each Purkinje cell in the cerebellum (Fig. 1a) receive synaptic input from more than 100,000 afferent parallel fibres that release glutamate as their neurotransmitter⁷. By using whole-cell patch-clamp recordings combined with confocal imaging in cerebellar slices, we recorded the parallel fibre-mediated electrical synaptic response at the level of the cell body and simultaneously monitored the associated changes in intradendritic Ca^{2+} concentration. Figure 1 shows an experiment in which two distinct synaptic inputs (Fig. 1a, regions 1 and 2) were activated consecutively with brief trains of afferent stimuli. Both inputs evoked the previously described¹⁰ ionotropic α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA)-receptor-mediated excitatory postsynaptic currents (EPSCs (Fig. 1c, e, insets). At the sites of synaptic contact, the stimulation produced a transient increase in the postsynaptic Ca^{2+} concentration that was confined to small dendritic regions (Fig. 1b, d). Unlike the previously reported local Ca^{2+} transients, which had a fast rising phase and a monotonic decay^{11,12}, this new postsynaptic Ca^{2+} signal had a more complex appearance consisting of two distinct peaks (Fig. 1c, e). The sequence of images

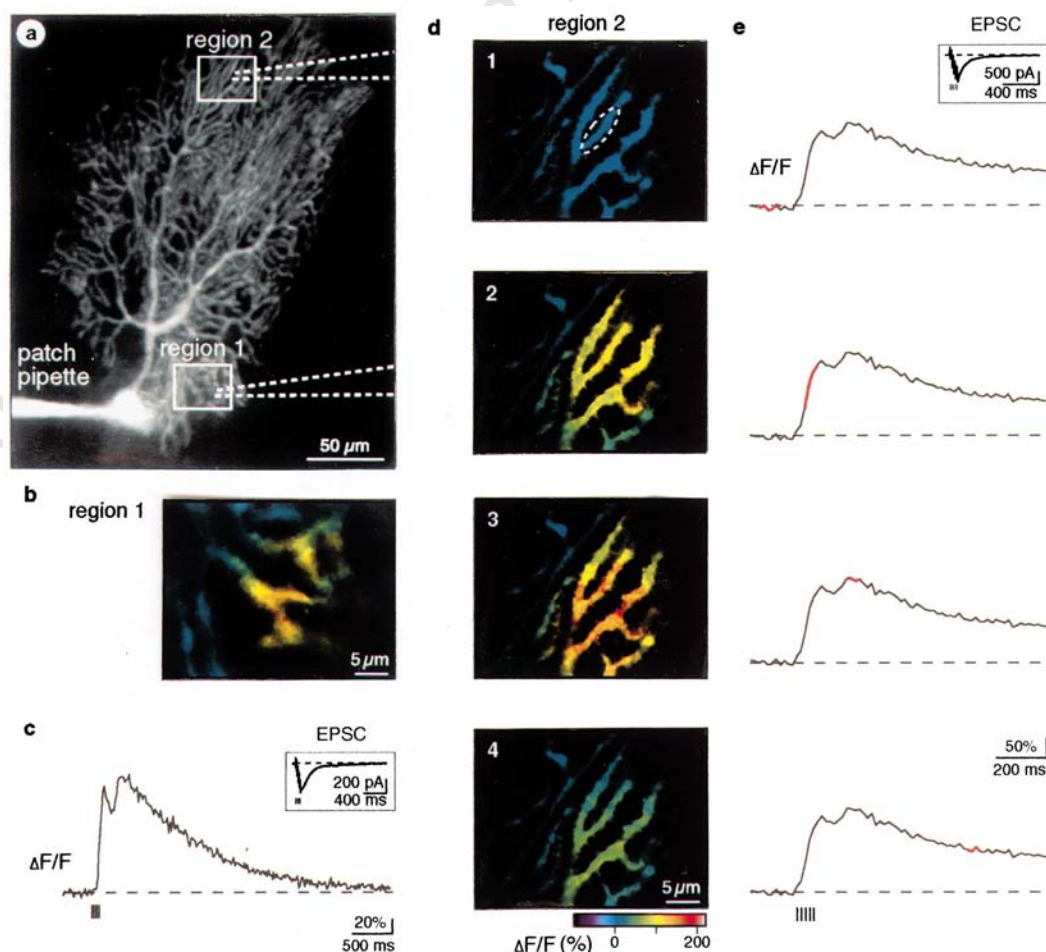


Figure 1 Local dendritic Ca^{2+} signals mediated by repetitive parallel fibre stimulation. **a**, Confocal image of a patch-clamped cerebellar Purkinje cell. The dendritic regions outlined by boxes are sites of stimulation shown at higher magnification in **b**, **d**. The positions of stimulation pipettes are indicated by dotted lines. **b**, Pseudocolour image of the relative increase in Ca^{2+} -dependent fluorescence ($\Delta F/F_0$) evoked by repetitive parallel fibre stimulation (5 stimuli, 50 Hz) in the proximal dendritic region 1.

The time course of the Ca^{2+} transient ($\Delta F/F_0$) and the simultaneously recorded EPSC (inset) are shown in **c**. Note the two peaks of the 'complex' synaptic Ca^{2+} transient. **d**, Sequence of pseudocolour images showing a transient increase in Ca^{2+} -dependent fluorescence in a distal portion of the dendritic tree (region 2). Images were taken at the periods indicated by red lines during the time course of the Ca^{2+} response shown in **e**. The corresponding EPSC is shown in the inset.

in Fig. 1d shows that the spatial dimensions of the postsynaptic Ca^{2+} response were almost constant throughout the entire duration of the signal. Similar synaptic Ca^{2+} signals were reliably obtained under our recording conditions (see below) and were detected in all Purkinje cells studied (65 inputs in total, 46 cells). These complex Ca^{2+} signals were obtained in all dendritic subfields, indicating that the receptors and channels responsible for their generation are uniformly distributed throughout the dendritic tree.

Pharmacological dissection of the complex synaptic Ca^{2+} response showed that 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX), an antagonist of AMPA receptors, blocked the first peak of the Ca^{2+} response (Fig. 2a). For simplicity we have termed the persistent, second component the delayed synaptic Ca^{2+} transient (DSCT). As expected, CNQX completely blocked the EPSCs¹⁰. In contrast, an antagonist of metabotropic glutamate receptors (mGluRs), (*R,S*)- α -methyl-4-carboxyphenylglycine (MCPG)^{2,3}, blocked the second component of the synaptic Ca^{2+} transient (Fig. 2a, c). The remaining 'early' synaptic Ca^{2+} transient (ESCT) had a time course that was identical to that observed with single-shock stimulation and is most probably caused by an AMPA-receptor-mediated local depolarization associated with Ca^{2+} entry¹¹. Remarkably, MCPG did not seem to affect the EPSCs at all (Fig. 2a). Similar pharmacological results were consistently obtained whenever tested (8 cells). These results indicate that the complex synaptic Ca^{2+} response may consist of just two components: an early one, mediated by AMPA receptors, and a later one that results from the activation of mGluRs. Consistent with this conclusion, the combined use of CNQX and MCPG effectively blocked the synaptic responses evoked in Purkinje cells by the repetitive stimulation of parallel fibres (Fig. 2b). What is probably the most important and surprising conclusion from these experiment

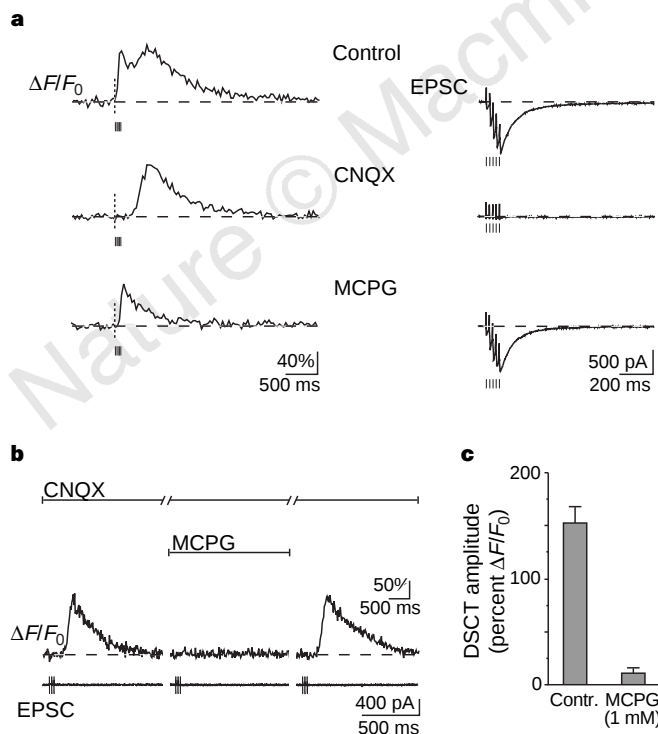


Figure 2 Pharmacological dissection of the complex synaptic Ca^{2+} signal. **a**, the AMPA-receptor antagonist CNQX (10 μM) completely blocked the parallel fibre-mediated EPSCs (right column), whereas the second Ca^{2+} component (shown by the change in Ca^{2+} -dependent fluorescence, left column) seemed unaffected. The mGluR antagonist MCPG (1 mM), applied after washout of CNQX, blocked the second component of the complex synaptic Ca^{2+} transient. **b**, MCPG (1 mM) reversibly blocked the delayed synaptic Ca^{2+} transients (DSCTs) obtained in the presence of CNQX (10 μM). Note the absence of EPSCs in these recording conditions (traces at the bottom). **c**, Summary of DSCT blockade by 1 mM MCPG ($n = 5$).

is that the new type of synaptic response consists solely of local dendritic Ca^{2+} signals. This contrasts with the slow synaptic depolarization involving activation of mGluRs that was first described in refs 13, 14. Those slow mGluR-mediated EPSCs can also be evoked in our recording conditions¹⁵, but seem to require significantly stronger afferent stimulation (H.T., J.E. and A.K., unpublished results).

The obvious signal cascade that may underlie the generation of DSCTs is the inositol-1,4,5-trisphosphate (InsP_3)-mediated Ca^{2+} release from the endoplasmic reticulum⁵ that follows the activation of mGluR1, a member of the group I metabotropic glutamate receptors that stimulate phosphatidylinositol hydrolysis⁴ and that are markedly expressed in the cerebellum^{16,17}. To test this hypothesis, we compared the complex synaptic Ca^{2+} response, under identical experimental conditions, in a control set of Purkinje cells and in cells that were dialysed with heparin, an antagonist of InsP_3 -mediated Ca^{2+} release from the endoplasmic reticulum in many preparations, including cerebellar Purkinje cells¹⁸ (Fig. 3a). In these experiments, which were optimized for a high DSCT:ESCT ratio signal (see Methods), heparin completely abolished the DSCT component. The application of cyclopiazonic acid (CPA), a compound known to block the ATP-driven uptake of Ca^{2+} into the endoplasmic reticulum and thus to mediate the depletion of these Ca^{2+} stores¹⁹, significantly reduced the DSCTs (Fig. 3b). Thus, three lines of evidence indicate that the DSCTs are mediated by the release of Ca^{2+} from intradendritic stores: the DSCTs are, first, blocked by intracellularly applied heparin, second, reduced by CPA, and third, not associated with any current through the plasma membrane (Fig. 2). Note that stable DSCTs were recorded, even in the absence of ESCTs, over extended periods of time (up to 20–30 min; Fig. 2b). This indicates that Purkinje cells, like CA1 hippocampal pyramidal cells, may not require Ca^{2+} entry through voltage-gated channels to refill their intracellular Ca^{2+} stores¹⁹.

DSCTs, unlike ESCTs¹¹, required repetitive stimulation (Fig. 4a–c). This is consistent with the hypothesis that, during single-shock stimulation, glutamate released from parallel fibre terminals will not reach significant concentrations at the perisynaptically located mGluRs²⁰. Alternatively, or in addition, repetitive stimulation may be necessary for the production of sufficiently high intracellular InsP_3 concentrations. In our experiments, a train of as few as two stimuli was sometimes sufficient to evoke DSCTs (Fig. 4b, c). The

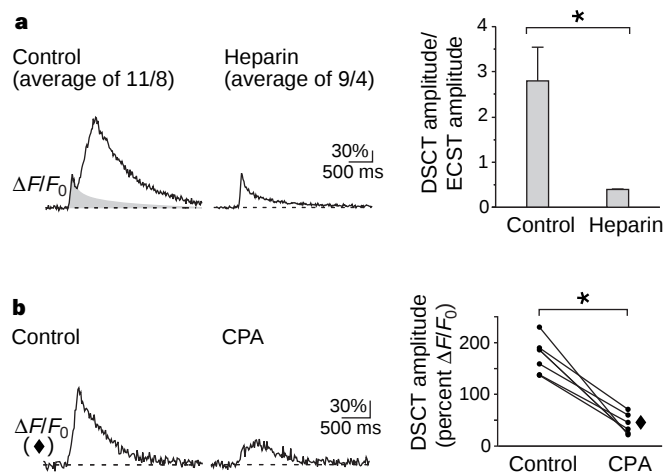


Figure 3 Intracellular mechanisms underlying the delayed synaptic Ca^{2+} transients. **a**, DSCTs were abolished by the intracellular application of the InsP_3 -receptor antagonist heparin (4 mg ml⁻¹). The traces represent averaged responses from control experiments ($n = 11$ inputs in total, 8 cells) and heparin experiments ($n = 9/4$). The grey shaded area in the left-hand trace indicates the contribution of the ESCT component. Results are summarized in the graph at the right. **b**, An antagonist of the endoplasmic reticulum Ca^{2+} -ATPase, CPA (20 μM), markedly reduced the DSCTs recorded in the presence of CNQX (10 μM). The graph (right) summarizes the results obtained in 6 cells. The diamond indicates the experiment illustrated on the left. Asterisk indicates $P < 0.001$, paired t-test.

threshold for the number of stimuli per train required for the generation of DSCTs, although not systematically investigated, appeared to be frequency-dependent (Fig. 4c), a finding that is also consistent with the requirement of transmitter accumulation in the synaptic cleft and/or the intracellular accumulation of InsP₃.

The onset of the ESCTs was steep (Fig. 4d), as expected for a process involving postsynaptic depolarization. The decay of both the ESCTs and the DSCTs was mono-exponential, with time constants of ~ 400 ms (Fig. 4d, e). Unlike synaptic transmission involving ionotropic receptors, the postsynaptic process of Ca^{2+} release from the dendritic endoplasmic reticulum underlying the DSCTs started with a considerable delay (Fig. 4e). Following this ‘release latency’ of ~ 200 ms, the Ca^{2+} concentration in the dendrites gradually built up, reaching its peak after another 260 ms (Fig. 4e, f).

To understand the function and the signal-integration capacity of this new type of synaptic response, it is important to know the precise size of the active postsynaptic domain. When the intensity of synaptic stimulation, and thus the number of active parallel fibres, was reduced, the postsynaptic Ca^{2+} -signalling domains changed their appearance from larger, compact spinodendritic domains (Fig. 1) into hotspots, with their 'centres' at the active spines. Figure 5a shows an example in which two adjacent active spines were detected (there may have been other active spines, outside the field of view). Spine 1 responded to repetitive stimulation by

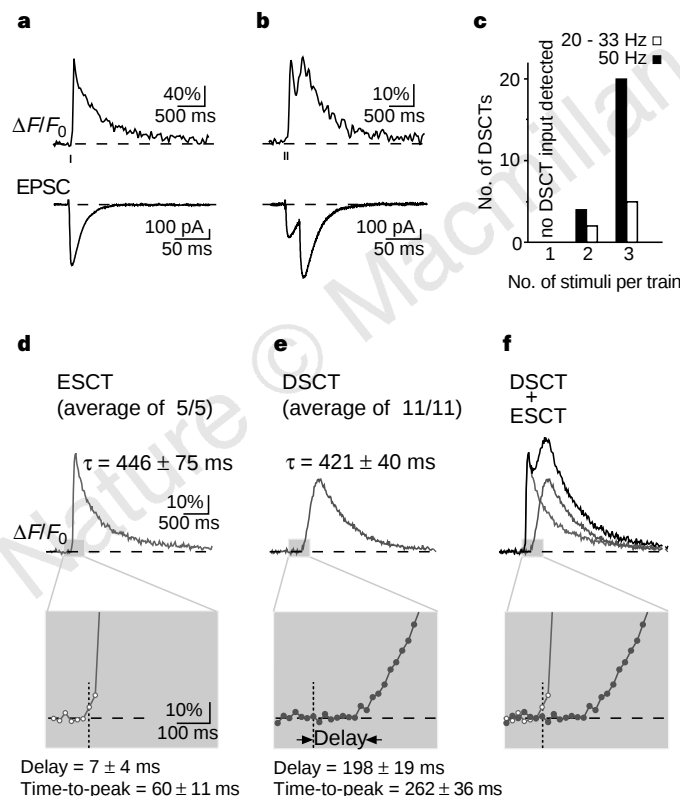


Figure 4 Mode of induction and time course of parallel-fibre-mediated synaptic Ca^{2+} transients. **a**, Pure ESCT (top) and the corresponding EPSC (bottom) evoked by single-shock stimulation. **b**, Complex synaptic Ca^{2+} response and the corresponding EPSC evoked by a 'train' of two stimuli in another cell. **c**, Summary of the minimal number of stimuli delivered to afferent parallel fibres needed to generate DSCTs in Purkinje cells. **d**, **e**, Average responses of **d**, ESCTs ($n = 5$ inputs/5 cells) and **e**, DSCTs ($n = 11/11$) recorded in the presence of 1 mM MCPG and 10 μM CNQX, respectively. The lower graphs show the initial portions of the responses at higher resolution. The mono-exponential decay constants, τ , the delay from the time of stimulation to the onset of the response and the duration from the onset to the peak of the response (time-to-peak) are given in the figure as means $\pm$ s.e.m. **f**, Calculated sum (black trace, upper panel) of the averaged ESCT (pale grey trace) and the averaged DSCT (dark grey trace).

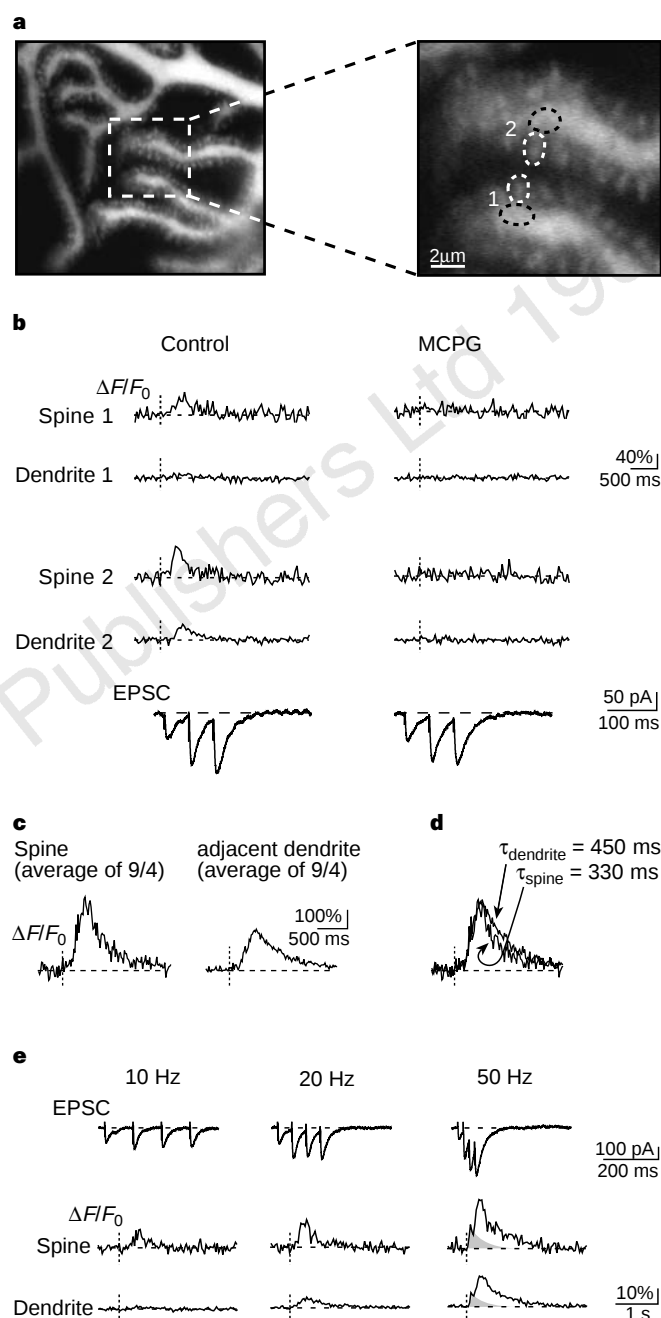


Figure 5 Identification of synaptic Ca^{2+} -release signals in spines and dendritic microdomains. **a**, Images of spiny dendrites of a Purkinje cell, indicating the sites of fluorescence measurements. **b**, Fluorescence measurements taken from the dendrites shown in **a**. In spine 1, the synaptic Ca^{2+} transient was restricted to an individual spine. Note the complete absence of any Ca^{2+} signal in the immediate dendritic vicinity. A large synaptic Ca^{2+} transient was detected in spine 2 and a smaller signal occurred in the subspine region. MCPG (1 mM) completely blocked all synaptic Ca^{2+} transients. The bottom traces illustrate the corresponding EPSCs evoked by a train of 3 stimuli (20 Hz). **c**, Averaged synaptic Ca^{2+} transients in pairs of spines and their adjacent dendrites (9 inputs/4 cells). **d**, Superposition of the normalized traces of **c**. Decay time constants are shown. **e**, Frequency dependence of the synaptic Ca^{2+} response in an individual spine and the adjacent dendrite (bottom traces; similar arrangement to that shown in **a**, **b**). The traces of the top row show the corresponding EPSCs. The frequency of stimulation is indicated at the top of each column. The presumed AMPA-receptor-mediated ESCT component (grey shaded area, right column) was identified on the basis of its short latency.

exhibiting a Ca^{2+} transient that was confined to the spine region, with no response in the adjacent subsynaptic compartment (Fig. 5b, top trace and e). Spine 2, instead, responded with a larger Ca^{2+} transient that was associated with a smaller transient in the immediate dendritic vicinity. These two types of 'minimal' postsynaptic Ca^{2+} signal had a release latency of ~ 200 ms and were abolished by MCPG (Fig. 5b), thus identifying them as DSCTs. The amplitudes of the Ca^{2+} signals were significantly larger, and the decay time courses of the Ca^{2+} transients considerably faster, in the spines than in the adjacent dendrites (Fig. 5b–d).

Figure 5e illustrates the frequency dependence of the postsynaptic Ca^{2+} response. At the lowest stimulation frequency (10 Hz), the DSCT was restricted to the spine. When the stimulation frequency was increased to 20 Hz, the amplitude of the DSCT in the spine increased and a small Ca^{2+} signal was also detected in the dendritic subsynaptic region (Fig. 5e, middle). When the stimulation frequency was increased to 50 Hz, the amplitude of the DSCTs increased further. Moreover, in parallel to the frequency-dependent build-up of the AMPA-receptor-mediated electrical response (Fig. 5e, top right), a short-latency ESCT was also detected in the spines and dendrites. Thus, even at the level of a single synaptic input, DSCTs seem to have the capacity to recruit, in a frequency-dependent manner, postsynaptic domains of increasing dimensions. A possible mechanism for this process is the frequency-dependent postsynaptic accumulation of InsP_3 . At high stimulation frequencies, the amount of InsP_3 will build up and invade neighbouring dendritic domains. Consistent with this model, InsP_3 receptors have been detected immunocytochemically in spines and dendrites of Purkinje cells^{21,22}.

Our results show that there are synaptic signals in spines and dendritic microdomains that require activation of mGluRs⁴ and mobilization of Ca^{2+} from the endoplasmic reticulum^{5,6}. An MCPG-sensitive postsynaptic Ca^{2+} component²³ and store-related Ca^{2+} release in spines²⁴ have also been found in hippocampal pyramidal neurons, indicating that DSCT-like responses may be a common form of synaptic signalling in the brain. There is, as yet, no evidence for a functional role of the synaptic Ca^{2+} -release signals during 'regular', low-frequency synaptic transmission. However, mutant mice lacking mGluR1 (refs 25, 26) or InsP_3 receptors²⁷ exhibit severe cerebellar defects, including motor coordination deficits and ataxia. The analysis of these mutants at the cellular level revealed a marked impairment of a characteristic cerebellar form of synaptic plasticity, the long-term depression (LTD) of transmission at the parallel fibre/Purkinje cell synapse^{25,26}. As induction of LTD requires both activation of mGluRs^{8,9} and a transient rise in the postsynaptic Ca^{2+} level²⁸, InsP_3 -mediated Ca^{2+} release seems to be an essential requirement^{27,29} for induction of LTD. These findings indicate that DSCTs may be involved in controlling activity-dependent synaptic plasticity of AMPA-receptor-based synapses in Purkinje cells. Our results suggest that this control may be exerted with a high degree of input specificity, even at the level of the individual synaptic contacts. □

Methods

Tissue preparation. Parasagittal cerebellar slices (200 μm) were prepared from 20–32-day-old CD-1 mice. The animals were decapitated following anaesthesia and the cerebella were rapidly removed and placed in artificial cerebrospinal fluid (ACSF) (0–2°C) containing (in mM): 125 NaCl, 2.5 KCl, 2 CaCl_2 , 1 MgCl_2 , 1.25 NaH_2PO_4 , 26 NaHCO_3 , 20 glucose, bubbled with 95% O_2 and 5% CO_2 . After cutting, slices were kept for 45 min at 37°C and then for up to 6 h at 20°C in ACSF³⁰.

Patch-clamp recordings. Somatic whole-cell recordings were obtained with an EPC9 amplifier (HEKA). We used 'Pulse' software (HEKA) for data acquisition. Pipettes (2.2–2.8 M Ω) were pulled from borosilicate glass (Hilgenberg) and coated with silicon (RTV 615, GE Silicons). The intracellular solution contained (in mM): 148 potassium gluconate, 10 HEPES, 10 NaCl, 0.5 MgCl_2 , 4 Mg-ATP, 0.4 $\text{Na}_3\text{-GTP}$, 0.1 Oregon Green BAPTA-1 (Molecular Probes), pH 7.3. In some experiments, a caesium-based solution, in which potassium gluconate was substituted by 134 mM caesium gluconate, was used. In these conditions, voltage control in

nominal voltage-clamp recordings, although not perfect, is improved and AMPA-receptor-mediated depolarization at remote dendritic sites is reduced¹¹. In the experiments including heparin (Fig. 3a), 4 mg ml⁻¹ heparin (Sigma) was included in the caesium-based internal solution. During the experiments, the slices were continuously perfused with ACSF (22–23°C) to which 10 μM bicuculline (Sigma) was added. Afferent stimulation was done by using a patch pipette filled with 1 M NaCl (1 M Ω resistance). The stimulation pipette was located at the molecular layer, 10–30 μm above or underneath the Purkinje-cell dendrite of choice. The stimulus pulse (100 μs duration) was 2–20 V.

Confocal fluorescence imaging. A confocal laser-scanning microscope (Odyssey, Noran) attached to an upright microscope (Zeiss Axioskop, $\times 63$ water immersion objective, NA 0.9) was used to acquire fluorescence images at 60 Hz in parallel to the whole-cell recordings¹¹. Ca^{2+} imaging was started at least 20 min after establishment of whole-cell configuration. Full-frame images were recorded on an optical disk (TQ2026F, Panasonic) and analysed off-line with the Image-1 software (Universal Image). The Ca^{2+} -dependent fluorescence signals from selected regions of interest were background-corrected and expressed as increases in fluorescence divided by the prestimulus fluorescence values ($\Delta F/F_0$) using Igor Pro software (Wavemetrics). The images shown in Figs 1, 5 are averages of 128 consecutive video frames.

Received 16 September; accepted 2 November 1998.

- Eilers, J. & Konnerth, A. Dendritic signal integration. *Curr. Opin. Neurobiol.* **7**, 385–390 (1997).
- Sladeczek, F., Pin, J.-P., Récasens, M., Bockaert, J. & Weiss, S. Glutamate stimulates inositol phosphate formation in striatal neurons. *Nature* **317**, 717–719 (1985).
- Masu, M., Tanabe, Y., Tsuchida, K., Shigemoto, R. & Nakanishi, S. Sequence and expression of a metabotropic glutamate receptor. *Nature* **349**, 760–765 (1991).
- Conn, P. J. & Pin, J. P. Pharmacology and functions of metabotropic glutamate receptors. *Annu. Rev. Pharmacol. Toxicol.* **37**, 205–237 (1997).
- Berridge, M. Inositol trisphosphate and calcium signalling. *Nature* **361**, 315–325 (1993).
- Furuichi, T., Kohda, K., Miyawaki, A. & Mikoshiba, K. Intracellular channels. *Curr. Opin. Neurobiol.* **4**, 294–303 (1994).
- Ito, M. *The Cerebellum and Neuronal Control* (Raven, New York, 1984).
- Kano, M. & Kato, M. Quisqualate receptors are specifically involved in cerebellar synaptic plasticity. *Nature* **325**, 276–279 (1987).
- Linden, D. J., Dickinson, M. H., Smeyne, M. & Connor, J. A. A long-term depression of AMPA currents in cultured cerebellar Purkinje neurons. *Neuron* **7**, 81–89 (1991).
- Konnerth, A., Llano, I. & Armstrong, C. M. Synaptic currents in cerebellar Purkinje cells. *Proc. Natl Acad. Sci. USA* **87**, 2662–2665 (1990).
- Eilers, J., Augustine, G. J. & Konnerth, A. Subthreshold synaptic Ca^{2+} signalling in fine dendrites and spines of cerebellar Purkinje neurons. *Nature* **373**, 155–158 (1995).
- Denk, W., Sugimori, M. & Llinás, R. Two types of calcium responses limited to single spines in cerebellar Purkinje cells. *Proc. Natl Acad. Sci. USA* **92**, 8279–8282 (1995).
- Batchelor, A. & Garthwaite, J. Novel synaptic potentials in cerebellar Purkinje cells: probable mediation by metabotropic glutamate receptors. *Neuropharmacol.* **32**, 11–20 (1993).
- Batchelor, A. & Garthwaite, J. Frequency detection and temporally dispersed synaptic signal association through a metabotropic glutamate receptor pathway. *Nature* **385**, 74–77 (1997).
- Eilers, J., Batchelor, A., Garthwaite, J. & Konnerth, A. Synaptic activation of metabotropic glutamate receptors produces dendritic calcium transients in cerebellar Purkinje cells. *Pflügers Arch.* **433**(Suppl.), 549 (1997).
- Martin, L. J., Blackstone, C. D., Huganir, R. L. & Price, D. L. Cellular localization of a metabotropic glutamate receptor in rat brain. *Neuron* **9**, 259–270 (1992).
- Shigemoto, R., Nakanishi, S. & Mizuno, N. Distribution of the mRNA for a metabotropic glutamate receptor (mGluR1) in the central nervous system: an *in situ* hybridization study in adult and developing rat. *J. Comp. Neurol.* **322**, 121–135 (1992).
- Khodakhah, K. & Armstrong, C. M. Inositol trisphosphate and ryanodine receptors share a common functional Ca^{2+} pool in cerebellar Purkinje neurons. *Biophys. J.* **73**, 3349–3357 (1997).
- Garaschuk, O., Yaari, Y. & Konnerth, A. Release and sequestration of calcium from ryanodine-sensitive stores in rat hippocampal neurons. *J. Physiol. (Lond.)* **502**, 13–30 (1997).
- Baude, A. *et al.* The metabotropic glutamate receptor (mGluR1 α) is concentrated at perisynaptic membrane of neuronal subpopulations as detected by immunogold reaction. *Neuron* **11**, 771–787 (1993).
- Satoh, T. *et al.* The inositol 1,4,5-trisphosphate receptor in cerebellar Purkinje cells: quantitative immunogold labeling reveals concentration in an ER subcompartment. *J. Cell Biol.* **111**, 615–624 (1990).
- Takei, K. *et al.* Ca^{2+} stores in Purkinje neurons: endoplasmic reticulum subcompartments demonstrated by the heterogeneous distribution of the InsP_3 receptor, Ca^{2+} -ATPase, and calsequestrin. *J. Neurosci.* **12**, 489–505 (1992).
- Frenguelli, B. G., Potter, B., Slater, N. T., Alford, S. & Collingridge, G. L. Metabotropic glutamate receptors and calcium signalling in dendrites of hippocampal CA1 neurons. *Neuropharmacol.* **32**, 1229–1237 (1993).
- Korkotian, E. & Segal, M. Fast confocal imaging of calcium released from stores in dendritic spines. *Eur. J. Neurosci.* **10**, 2076–2084 (1998).
- Aiba, A. *et al.* Deficient cerebellar long-term depression and impaired motor learning in mGluR1 mutant mice. *Cell* **79**, 377–388 (1994).
- Conquet, F. *et al.* Motor deficit and impairment of synaptic plasticity in mice lacking mGluR1. *Nature* **372**, 237–243 (1994).
- Inoue, T., Kato, K., Kohda, K. & Mikoshiba, K. Type 1 inositol 1,4,5-trisphosphate receptor is required for induction of long-term depression in cerebellar Purkinje neurons. *J. Neurosci.* **18**, 5366–5373 (1998).
- Konnerth, A., Dreessen, J. & Augustine, G. J. Brief dendritic calcium signals initiate long-lasting synaptic depression in cerebellar Purkinje cells. *Proc. Natl Acad. Sci. USA* **89**, 7051–7055 (1992).
- Khodakhah, K. & Armstrong, C. M. Induction of long-term depression and rebound potentiation by inositol trisphosphate in cerebellar Purkinje neurons. *Proc. Natl Acad. Sci. USA* **94**, 14009–14014 (1997).
- Edwards, F., Konnerth, A., Sakmann, B. & Takahashi, T. A thin slice preparation for patch clamp recordings from neurons of the mammalian central nervous system. *Pflügers Arch.* **414**, 600–612 (1989).

Acknowledgements. We thank O. Garaschuk for comments on the manuscript and R. Trautmann and E. Eilers for technical help. This work was supported by grants from the DFG and the HFSP.

Correspondence and requests for materials should be addressed to A.K. (e-mail: akonnerth@med-rz.uni-sb.de).